Audio Stream Bookmarking with a Wristband Controller:
Exploring the Role of Explicit Commands in an Implicit Control Loop

Jih-Shiang Chang, Joanna McGrenere, Karon E. MacLean
Department of Computer Science, The University of British Columbia
Vancouver BC V6T 1Z4
{gordonjc, joanna, maclean}@cs.ubc.ca

ABSTRACT
This project places itself as a preliminary and informal evaluation to explicit commands and a proposed implicit control loop. It focuses on the design methods for explicit commands and the relationship between implicit and explicit control channels. Attentional bookmarking (interruption driven) for spoken audio streams (e.g., audio books, podcasts) are chosen as a sample use case. The full project is comprised of information gathering of the use case, brainstorming and designing a controller prototype, and a preliminary field evaluation of the controller. Several design implications for explicit commands and implicit control loops are suggested.

Author Keywords
Podcasts, audio books, bookmark retrieval, attentional bookmarking, physiological signals, galvanic skin response, interruption, explicit commands, explicit path, implicit path, implicit control channel, wearable devices.

INTRODUCTION
Communication between users and modern user interfaces often requires users' full attention and a solid understanding of the commands or the controls, especially via the visual and audio channels. It is often mandatory for users to alter or receive system status explicitly. This kind of explicit information exchange often requires a high level of user attention. Also, multitasking has become a common occurrence in daily life. People constantly switch between multiple tasks throughout the day. The cost of task-switching can sometimes be high, which results in an increase in frustration and inefficiency [1].

In order to mitigate this problem, interaction designers and technology innovators have devoted themselves to coming up with better designs and technology. A great amount of research has focused on "implicit interactions", trying to understand users' preferences and behaviour in order to properly predict the demands that users have at appropriate times. One way to achieve this goal is by studying explicit commands in the users' command history or from other users, which is extensively developed in areas of user-adaptive systems, machine learning, and artificial intelligence. On the other hand, researchers have studied implicit control channels for interaction, such as eye-tracking, physiological signals, and unconscious motion. These implicit control channels require a very low level of attention.

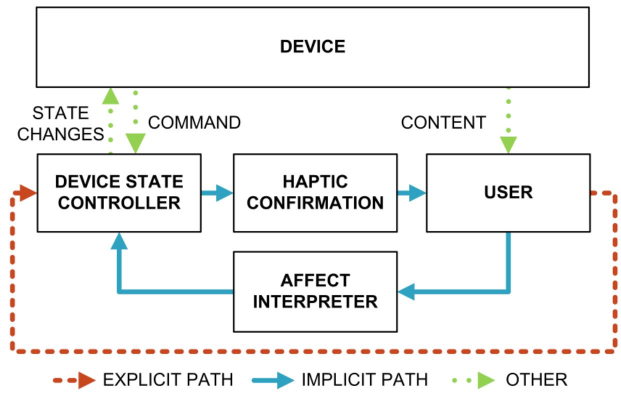

Figure 1. Proposed implicit control loop structure.

There are limitations to implicit control channels. It is hard predict with 100% accuracy what users want without explicit commands. Also, users may prefer to possess a high level of control to override system commands at their own will depending on the specific context [2]. Thus, we believe that a traditional explicit control channel is still desirable as a supportive path within implicit interaction processes. As far as we are concerned, explicit paths to system input can be utilized to 1) retrieve information collected from implicit paths 2) rapidly correct or tune system behavior 3) train and improve the system's affect modeling and decision making strategy.

In our previous work [3], we proposed an implicit control paradigm for device interaction. In the paradigm, shown in Figure 1, we minimize attentional requirements by analyzing and capturing continuous affective states through biometric sensors as an implicit path to trigger system change, and use the haptic channel to provide feedback of system change unobtrusively. In ongoing work [4, 5], the role of haptic feedback (the "haptic confirmation" box in Figure 1) is being investigated, and algorithms for affect interpreter are being evaluated [2]. The focus of this project is to explore the role of explicit path in this diagram.

We are interested in how the explicit path should be designed in an implicit control loop, in order to maximize user satisfaction. We approach this problem by exploring a similar context to our previous work [3], which involves listening to spoken audio streams (e.g., audio books, podcasts, radio drama) when multi-tasking. Consider the following scenario:

A man is listening to a podcast attentively (task 1) while walking on the street (task 2). Suddenly a friend bumps into him unexpectedly. He quickly takes off the headphones and starts chatting with the friend (task 3). After a couple of minutes, the conversation ends, and he puts his headphones back on and tries to seek the position where he has left off.

Usually it requires significant effort and/or time to rewind the audio player to the exact desired position. In this scenario, many users are likely to end up neglecting the missing pieces of the podcast (undershooting) or repeat listening to the same point (overshooting). This interruption-driven scenario is chosen for exploring the roles of implicit and explicit control channels as it involves two steps:

1. *When the user diverts his attention immediately to the interruption, an implicit control is needed to facilitate the change by acknowledging and recording the position where the user left off.*
2. *When the user returns his attention back to the podcast, an explicit path to control the system is required. The command should support retrieving the position recorded from the implicit channel.*

These occasions construct a suitable scenario for investigating the design issues and employment of explicit paths in the proposed interaction control loop diagram. Thus, our targeted use case is defined as attentional bookmarking (interruption driven) for spoken audio streams.

In this project, we first conduct a thorough interaction design process with respect to the use case of spoken audio streams. The purpose of interaction design is to refine a valuable use case and to provide a functional prototype that is as suitable for investigating explicit and implicit control channels as possible. Then, through a preliminary study, we explore and analyze the role of explicit commands for systems with implicit control loops.

This project begins with information gathering (summarized in `Related Work and Survey`), followed by design approach and rationale (detailed in `Designing the Controller`), prototyping (illustrated in `Controller Prototype`), and preliminary evaluation (described in `Preliminary Field Evaluation`). We report our findings, discussion, and future work based on our learning throughout this project (stated in `Discussion` and `Conclusion and Future Work`).

RELATED WORK AND SURVEY

Our literature review covers explicit and implicit interaction to understand the trends in related interaction design work. In addition, we survey existing media players for audio books and summarize their common features as they are essential to evaluating our use case. We also touch on bookmarking features for media streams. Lastly we talk about work related to interruption detection for attentional bookmarking.

Explicit Commands and Implicit Interaction

Traditional input methods, such as the WIMP paradigm (Window, Icon, Menu, Pointing device) require users' full visual attention to perform a command. Recently, designers have produced sophisticated interaction tools such as Apple Magic Mouse [6] or have created intuitive multi-touch gestures (e.g. Synaptics Gesture Suite [7]), to perform more tasks. Irani proposes mixing input modes to enhance interaction [8]. Weiser and Brown propose calm technology [9], a well-praised concept in interaction design that device interfaces can "*move easily from the periphery of our attention, to the center, and back*". One way to achieve this goal is to use haptics, the sense of touch, as a means for background display [4]. As technology becomes more mobile-oriented, wearable haptic or non-haptic devices is an emerging trend [10].

Hand gestures can be one of the most natural ways to give explicit commands [11, 12], in both desktop (e.g., [13]) and mobile environments (e.g., [14]). Explicit interaction doesn't always involve direct contact with the device interface. It has the potential to immerse the user and real-world artifacts as well. For example, SixthSense [15] tries to connect virtual objects and real-life objects as a family to facilitate simple interaction. Harrison et al. propose a muscle-computer interface by collecting Electromyography signals [16, 17]. From the literature, we conclude that there is an advantage to applying gestures on wearable devices for explicit interactions.

More and more products and research integrate traditional explicit interaction paths and various sensory data as implicit input to a platform for better human-computer interaction. These include tracking movement (e.g. Wii, iPhone), physiological signals (e.g. [16-20], and the Wii vitality sensor3), accelerometers and GPS (e.g., [21, 22]), or a combination of the above (e.g., [23]).

Ju et al. [24] present a framework for implicit interactions. They propose that interaction designers should study the transitions between explicit commands and implicit interactions. Note that implicit interactions mean the ability of a system to communicate and perform actions without explicit input from users; this includes studying users' behaviors continuously or analyzing data from implicit control channels. In Ju's paper, implicit control channels are not fully considered and investigated in interaction design, while our work focuses directly on exploring explicit commands in an implicit control loop.

Survey on the Use Case

Due to the popularity of portable media players and digital audio files, the number of podcasts and audio books listeners has rapidly increased for the past ten years. As a result, software applications have been designed specifically for audio books, mostly for mobile devices (e.g., [25]). These players have unique features over traditional music players, such as chapter based navigation, one touch buttons for jumping backward or forward 15 seconds, and bookmarking with notes. iTunes also has distinct categories for podcast and audio books download and management. These features require both visual cues and focused user attention, which is undesirable in mobile contexts. One of the guidelines we followed when designing our prototype was to minimize visual attentional requirements as much as possible.

Navigating through media files or negotiating bookmarks are what we presume to be the most frequent tasks in our use case. Research on effective video stream browsing is common. A study from Fels et al. creates bookmarks with different colors to provide a better interaction for video stream browsing [26]. While typical data streaming for a media player is visualized on a screen, Snibbe and MacLean propose using haptic metaphors for eyes-free digital media consumption [27]. Little research has focused on retrieving bookmarks for audio books without visual cues, nor on design methods for retrieving auto-generated bookmarks. We believe applying haptics is an effective strategy for our use case (Figure 1). However, haptic feedback is not included in our prototype system at this time. We use audio feedback instead due to its rapid implementation.

Implicit Interruption Detection

External interruptions often correspond with an Orienting Response (OR), which is an immediate physiological reaction to a change in surroundings. OR is recognized widely at the start of attentional shifts [28]. Galvanic skin response (GSR), the electrical conductance of the skin, is recognized as a valid measurement for ORs [29, 30]. StartleCam monitors a person's skin conductance to detect whether that person is startled [31]. It uses a wearable video camera to save (bookmark) the digital images as an offline image-based diary. Gazemarks is another application that bookmarks the location of gazes on the screen and highlights this area for you when you return from other windows or from an external interruption [32].

In our previous studies [3], detection of verbal external interruptions through GSR signals reached peak accuracy of 84% while participants listened to audio books. This is proof of the technical feasibility of attentional bookmarking (interruption driven). However, a preliminary study on collecting GSR signals in walking and running shows that GSR signals are sufficiently sensitive to body movement that the current algorithm fails to detect ORs. We still envision detection can be successful in mobile contexts via noise filtration of GSR signals. Applying context awareness, including analysis of gait through accelerometers and GPS locations, is our approach. In our project, we have assumed that the technology will be feasible in near future.

DESIGNING THE CONTROLLER

We aim to explore and define our initial use case more precisely. We interviewed typical users to characterize user profiles. We refined our use case and created feature requirements for a spoken audio stream controller. Based on interviews and a systematic review of commercial products, we devised ideas for designing the controller. This work is necessary as we want to explore explicit commands in a well defined use case with a proper prototype.

Informal Interviews with Colleagues

We conducted an informal interview with heavy consumers of podcasts or audio books to understand their listening experience in the context of their lives, to identify interruption cases, and to discuss optimal interaction techniques for implicit and explicit control paths when listening.

In this semi-structured interview, we first collected demographic information and built a listener profile, asking respondents how they listen to audio books and/or podcasts, what they listen to, when do they do so, and why. Subjects were asked to talk about possible interruption cases in their personal listening experience. We also discussed their frequency, importance, and how they would react to four sample interruption situations:

1. *User gets on the bus and seeks a vacant seat.*
2. *User sits on a couch at home, and a family member or roommate enters the room.*
3. *User drives on the road and sees an accident ahead.*
4. *User does chores and has to pause work when the phone rings.*

Afterwards we talked about audio players and their functions, to discover general design conventions and user comments. The bookmarking feature was specifically addressed. Lastly we introduced the implicit control loop diagram (Figure 1) in the context of attentional bookmarking during audio book listening [3], in which the system automatically bookmarks the position of audio files such that the user can divert his/her attention to interruptions, then resume listening from the point he/she is interrupted. Interaction techniques for bookmarks confirmation, notification, and navigation were discussed. Subjects were asked to picture an ideal yet realistic system on the basis of promising technology. Each interview was estimated to last for 40-60 minutes.

Interview Results

Four university students volunteered to take part in the interview, ages ranging from 25-32 years old. Based on the verbosity of the participants, the interviews lasted up to 75 minutes. All subjects were identified as heavy and frequent listeners of spoken audio stream (audio book and/or podcast) listeners. They listen to content 40 minutes to 2 hours per day. They typically listen to spoken audio streams when *commuting* (e.g. walking from one location to another, busing, or driving), *exercising* (e.g. biking), *street-wandering* (Subject 2 only), or *working in an office* (Subject 4 only). They listen to spoken audio streams under above situations mainly because they want to make full use of their time. The spoken audio streams subjects listen to can be categorized into *designed lessons* (e.g. for language learning, knowledge learning), *entertainment* (e.g. talk shows, storytelling), *news*, or *inspiring, interesting topics* (e.g. speeches, radio programs). iPod, iPhone, or traditional MP3 players are the typical devices for listening, mostly coupled with a pair of headphones. Subjects tend not to have vision required interaction with the device interface because it is cumbersome. They choose the desired audio file(s) and stick to that particular file(s). Subject 3 (S3) reported that he splits the audio file into 5-minute long chunks so that he can always recognize his progress. S4 reported that she sometimes treats the audio stream as "background sound" and neglects the content when an interruption occurs.

Regarding interruption cases, sample situations were generally acceptable. Cases 2, 3, and 4 were identified as less relevant to their own experience. Additional cases were suggested and categorized as follows:

5. *User walks on the street and someone asks for directions.*
6. *User walks on the street and strangers walk by.*
7. *User works in an office and someone approaches to start a conversation.*
8. *User walks on the street and the phone starts ringing*
9. *User walks to an ATM or a kiosk at train station to use it.*

Current functions on their players are acceptable. All subjects desire the player to have a play-and-pause button, rewinding functionality, and the ability to memorize the location of pauses. Interfaces were highly praised among subjects to be intuitive and easily accessible. Interesting enough, S1 and S3 have never heard of bookmarking features (except for memorizing the last played location, if it qualifies). S3 mentioned that he would love to bookmark points of interest; S2 has heard of it but has never had a chance to use it; S4 has bookmarking feature on his device but is reluctant to use it because "*it is too much of a hassle to seek for the bookmarks if I need to take the player out (of my pocket)*".

Given that attentional bookmarking is a promising technology, strong common reactions were 1) confirmation is not needed; it is frustrating 2) a system placing a bookmark and continuing playback is favored over a system pausing the player on interruption. Continued playback was considered more natural and less annoying. There was a diversity of opinions on notification methods and post-interruption navigation. S1 preferred audio notification over haptics, while the rest, had the opposite opinion. S3 suspected that vibrating notifications will be distracting to both the interruption and enjoying content. On the topic of bookmark navigation, S4 hoped to have a visual interaction. S2 proposed a force feedback on the wheel for bookmark navigation; other subjects preferred a "simple" way for retrieving bookmark locations but could not propose a clear solution. S1 mentioned that it would be great if the system recognizes and recommends a resume command to the user whenever the interruption ends. S2 and S3 were concerned about the number and accuracy of system generated bookmarks; S3 indicated that if the system "screwed up", he would rather turn off the bookmarking feature (implicit path). Subjects mentioned that playback should resume a few seconds advanced of the actual bookmark, in response to [3]. Overall, subjects would be happy to see the proposed technology realized, but only if the interaction would be "natural". In terms of the location of the interaction interface, three subjects held a strong preference for something wearable, but definitely not on the head.

Refining the Use Case

Based on the results of the interviews, a new representative use case emerged:

> *Users want to concentrate on the content of the spoken audio streams, either podcasts or audio books, while doing other physical activities that do not require a high cognitive workload (e.g., commuting, exercising, doing chores). Usually this happens in a mobile context, where a portable player is essential. Users care about both their primary activity and the audio streams; they are parallel tasks. The interruption detection applies when an external interruption occurs; the timing of the interruption is unexpected, and the user's initial reaction is to respond to it immediately (orienting response).*

The new use case emerged in order to have a targeted end user, to prepare for proper prototype design, and to target a precise evaluation process. This use case appears to capture a common scenario, based on our interview results, which makes it strongly suitable as a test case to explore how explicit interaction should be designed in an implicit control loop.

Feature Requirements and Discussion

Since users listen to spoken audio streams mostly in mobile contexts, the controller device for explicit interaction should be portable, easily accessible, and possibly support simple hand-initiated gestures. Common audio players (i.e. phones or MP3 players) are normally kept in pockets or bags and are not ideal for quick and simple explicit interaction. We propose that the ideal controller is not only portable but also wearable. The explicit control channel should be simple, natural, yet easily distinguishable from common body movements.

We conclude the controller should have at least the following features:

a. *Play / Pause*
b. *Rewind a small number of seconds (15 seconds is a common default length in current audio book players)*
c. *Manually place a bookmark*

d. *Auto-bookmark generation based on external interruption detection*

e. *Bookmark navigation*

f. *Volume control*

Design Alternatives for the Controller

We brainstormed several designs, all of which are wearable and easy accessible. We studied and tested the feasibility of each design idea. Since the attentional cost of visual interaction is normally high, we avoided designing a controller that requires user's visual attention in this project.

Controller on the Wrist

Several research and prototypes on the market implement a wrist controller [33, 34], since the user can effortlessly perform a command, most of the time *eyes-free*. In our pilot studies, a designed tap-based wristband controller was capable of sensing taps even when the controller was hidden beneath sleeves, which brought flexibility when initiating commands. A watch-like controller with rotatable outer bezel (similar to iPod's scroll wheel) is a design alternative for bookmark navigation. The bezel can be mechanically rotatable or touch-based. Due to time frame of the project, we only tested it with a touch-based rim (a round shape Softpot Rotary Potentiometer) and found out that it is susceptible to erroneous touches. A wrist-mounted touch screen (such as the new iNano [35]) is good for complex interaction but it requires visual attention.

Controller Integrated with Earphones

Several commercial products were successful on the market due to expressive yet easy-to-access interfaces on earphone wires (e.g., NOKIA WH-701). Apple's earphones with remote control and microphone [36] have a convenient set of buttons for adjusting volume and simple playback functions. The limitation of this kind of controller is that it is hard to perform complex yet intuitive interactions; it is also hard to grasp on to and perform a sophisticated action when the wire is soft. In our pilot studies, subjects reported that, when clicking on or making a sliding gesture on the earplugs, the noise the finger produces is irritating. In addition, fiddling with the earplug unnaturally is often unpleasant and embarrassing. Making earplugs the controller is therefore not ideal.

Controller with Hand Gestures

Previous effort has been expended recognizing hand gestures as interaction techniques [37-39], applying flex sensors and accelerometers on a glove to detect position and movement of the fingers. Nonetheless, having a glove worn on one's hand all the time reduces the ability of performing daily tasks such as grabbing the pole on a bus, washing dishes, or using a mobile phone.

In our brainstorming, several gestures were proposed to symbolize interactions in our bookmarking scenario. For example, having one's left hand curved as if it is holding a hamburger, and the other hand inserting into it imitates placing a physical bookmark; finger tips brushing on the palm of another hand mimics rewinding or fast-forwarding the media; an "ok" sign with the hand shows confirmation; tapping the fingers against each other means play/pause. It is questionable whether these gestures are distinguishable and are recognizable among other unconsciously generated gestures. However, these intuitive interactions show potential and require further study and development.

Others

A Sleeve-like wearable display designed by Pan et al. [40] is capable of providing directional haptic feedback, but it is harder to put on and take off than the alternatives. Squeezing an object mounted on the body is another way to facilitate interaction. Naturalness and representativeness of the metaphor in this case, however, are unsolved issues.

Chosen Design

Appropriateness, naturalness, and convenience of the controller were considered in order to find the best fit to our use case and goal. Based on the simplicity and the scope of the project, a tap-based wristband controller was chosen as the prototype. Proposed functionality and preliminary evaluation are described in the following sections.

Figure 2. Upper figure: components of the prototype. Lower left figure: snapshot of the prototype. Lower right figure: snapshot of the prototype being worn on a person's wrist.

CONTROLLER PROTOTYPE

The ideal system employs a wristband controller integrated with wireless GSR sensor [41]. The media player is either embedded on the controller or separately an individual player (e.g. iPhone) that communicates with the controller wirelessly. A pair of wired or wireless earphones conveys the audio stream from the player.

In our prototype, we use Arduino LilyPad microprocessor to connect all electronic components on the controller. The LilyPad connects to a laptop. Thought Technology's ProComp Infiniti® physiology-measurement hardware system [42] is used to collect GSR signals from the subject during our evaluation study for future analysis. The GSR sensor is placed on the finger tips of index and middle fingers of the participant, connected to the ProComp Infiniti® encoder, then to the laptop. A customized program on the laptop controls Windows Media Player with the functions described in Table 1. A pair of earphones is connected to the laptop.

Since subjects from the interview had diverse opinions on the optimal modality for bookmark navigation, we propose a thorough study on modalities (audio feedback, haptic feedback, and visual feedback) in our defined use case. Due to the time constraint of the project, in our prototype, the haptic feedback component is illustrated but not implemented; we use audio feedback for all kinds of system notification. Each type of explicit (e.g. pause) and implicit control channels (e.g. system detects an interruption and generates a bookmark) is assigned to a unique symbolic piece of audio feedback.

We designed the functionality based on the feature requirements defined above. Volume control is assumed to be integrated into the earphones' wires; hence it is not included in the prototype. The functionality implemented in the prototype is listed in Table 1.

Controller Components

The prototype was designed iteratively. The final version is illustrated in Figure 2. The prototype contains the following components:

Velcro wrist strap: A strap to hold all components together. The strap can function as a wristband by folding it round the wrist and fastening it to itself.

C2 Tactor: A vibrotactile transducer [43] to detect finger taps. It can detect a combination of several taps in a row [2]. The tactor also has the ability to vibrate and provide haptic feedback.

Bookmark navigation switch: A surface-mounted 3-way switch allows 3 modes of interaction. The switch is made of a piece of plastic and allows the user to navigate using up and down selections (by pushing the plastic piece to one direction or another), and pressing the plastic piece inwards.

Recovery button: A discrete push button; in our prototype, a LilyPad Button Board.

Name	**Description**
Pause or play	Single tap on the tactor to toggle between pause and play.
Place a bookmark	Double tap on the tactor to manually place a bookmark.
Forward or Rewind (small amount)	Slide on the pressure sensor (small quantity of movements on the wrist strap).
Forward or Rewind (large amount)	Slide on the pressure sensor (large quantity of movements, the fingers slides on both the wrist strap and the fabric).
Undo	Press on the recovery button to resume to previous location.
Go to previous bookmark	Use left selection on bookmark navigation switch.
Go to next bookmark	Use right selection on bookmark navigation switch.

Table 1. The function list for the wristband controller prototype. For forward and rewind, sliding to the right direction indicates forward, and to the left as rewind.

Fabric: A silk ribbon fabric to differentiate the texture of Velcro wrist strap.

Force sensing resistors: A pressure based sensor to detect finger movement such as sliding and long pressing.

PRELIMINARY FIELD EVALUATION

The purpose of conducting a field evaluation is simple. We simulated the use case in our study with the prototype to explore the role of explicit commands in designed implicit control loop. Subjects walk a familiar route so that they are immersed in their surroundings, and must face in-context interruptions and provide realist reactions. The evaluation is to have them test the scenario with given prototype, so that they have a deeper understanding of the explicit and implicit control loop, and provide the most relevant feedback possible on the proposed interaction. As we stated in previous sections, our interruption detection algorithm excels only when users are regularly motionless. Since the field evaluation is in a mobile context, we chose the Wizard of Oz evaluation technique for implicit interruption detection; in other words, a researcher watched the subject and determined when an interruption occurred. We believe that subjects are still able to experience how explicit path and implicit path contribute to the whole interaction loop, resulting in valid and fruitful feedback.

Design

The evaluation was designed as an outdoor commuting trip. The trip included a 20-minute walk and a 3-minute bus ride. Subjects were asked to bring 1) two interesting podcasts or audio books that they have not yet listened to, each at least 30 minutes in length, and 2) a pair of headphones they commonly use. The study was about 1 to 1.5 hours long, including: *introduction and device set-up*, a *practice session*, an *evaluation session*, and *post-session interview*. The experiment procedure is provided below. Transcript of the experiment procedure can be found in Appendix A.

Introduction and Device Set-up

Subjects were first introduced to the purpose of the study and the purpose of each device. They understood it is a preliminary study, and that the prototype is not a fully developed system. They then were asked to wear three things: 1) GSR sensors on the middle and index finger tips of their left hand, 2) headphones connected to an extension cable which connects to a laptop, 3) the wristband controller prototype on their left hand. They were asked not to manipulate with the GSR sensors at anytime. The subjects were then taught clearly about each system function and how to interact with it if applicable, along with an explanation of system feedback. Subjects were asked to experiment with the prototype, practicing the functions, while the researcher tested the signal and confirmed the subjects could perform each command successfully.

Subjects were explicitly told that both system-generated bookmarks and explicit commands would be simulated using Wizard of Oz, where the researcher observes and simulates the optimal behaviour of the actual system as it would perform. Heuristics to determine valid external interruptions (which generate auto-bookmarks) were discussed with each subject. The heuristics and actual GSR responses during the practice and evaluation sessions helped the researcher better understood when to “generate” a bookmark. Appendix B shows the sample list of heuristics verified and commented upon by subjects.

Practice Session

During both practice session and the evaluation session, the researcher walked beside and slightly behind the subject, holding a laptop which connected the sensors and headphones, to observe the subject and his/her surroundings. Subjects were asked to act as natural as possible. They had full freedom of what they can do (e.g. stop by a coffee shop to buy coffee, respond to phone calls, talk to someone). The only way to pause or to terminate the study was either the subject or the researcher performs a hand signal (an “X” signal, crossing arms in front of the chest). Subjects could terminate the study at their own will at anytime.

In the practice session, subjects walked to a nearby coffee shop within 5 minutes. They practiced the usage of the controller and the setting of the study. They had the chance to get a sense of how it actually feels to participate in the study.

Evaluation Session

In the evaluation session, subjects spent about 5 minutes walking to a bus stop, getting on the next bus, and getting off after two stops (the terminal bus stop). They then walked to the lab room for about 15 minutes. A camera-man was designated and introduced to the first subject. The video recording focused on the interaction on the subject’s wrist and surrounding sounds. We removed the camera-man from the scenario for the rest of the subjects to keep the study as natural as possible.

Three artificial interruptions were designed for the evaluation session: 1) A phone call from a friend 2) A stranger asking for directions 3) Bumping into a lab-mate or a friend. The detailed set-up for these interruptions is described in Appendix C. In addition, there were two designed natural interruptions: 4) Subject gets on the bus and looks for a seat 5) Subject walks by a construction site in progress. These interruptions were chosen mainly based on our previous interview results, described in `Designing the Controller`.

Post-session Interview

After the evaluation session, subjects took the GSR sensor and earphones off. A largely unstructured interview with the researcher lasted for around 25 minutes. Subjects were free to provide feedback on controller usage, give opinions on the flow of bookmark retrieval, and make general comments on the study. Each occurrence of designed or unexpected interruptions were discussed and evaluated. Finally the subjects removed the wristband controller.

Results

Four university students participated in the study, age ranging from 19 to 35 years old. One of the subjects also participated in the informal interview study described above. We first collect and analyze data on an aggregated level, and then discuss specific feedback from individual subjects.

Across four subjects and five designed interruptions, 20 intentionally designed interruptions are expected. However, only 12 designed interruptions are successfully performed and recognized by the subjects as valid interruptions. We categorize the user actions following these 12 valid designed interruptions: 5 bookmark navigations, 3 rewinding, and 4 pauses and resumes. There are also 18 other undesigned external interruptions or unknown factors that resulted in rewinding events or bookmark navigation. We also collected the user actions following these 18 interruptions: 14 bookmark navigations and 4 rewinds.

In general, subjects found the prototype and the interaction techniques to be well designed. The wrist was reported by all subjects as the proper location for explicit input in the context. S3 had to take off her watch during the study in order to wear the prototype. S2 suggested that the controller should be waterproof and must be wireless; he found the wires too distracting during the study. S4 wished that the GSR sensors could be put on other body parts aside from the finger-tips. All subjects found the navigation switch very neat, though S4 was concerned that the switch may be easily broken. The “undo” function was well appreciated (S1, S2, S3), though it was seldom used during the study.

S1 and S2 mentioned the ease of using a single tap to play and pause.

The audio feedback performance was mediocre. Subjects felt the audio feedback "beep" was a little bit annoying and the corresponding audio sounds were not well designed. S1 did not notice most of the audio feedback when he was interrupted. S1 proposed that tactile feedback might produce a clearer message. S2 suggested producing a "bump" through haptics when the user passes a bookmark while rewinding.

S3 and S4 reported that the bookmarking idea is good because "(S4) *by the time you think of placing a bookmark or pausing (the player), it's normally too late*"; S1 and S2 preferred pausing the player over utilizing the bookmarks. S1 wished that the player could automatically pause when he removes the earphone(s) from his ear. S2 and S3 mentioned that the bookmarks can be temporary and be removed after a while.

All subjects felt comfortable having both implicit and explicit paths to control the device, mainly because they had full control of the player and knew how it operated. They considered the interaction paths to be complementary. Training and trust issues were commonly brought up in the interviews. Subjects reported that they learned to trust the bookmarks after a while. S3 mentioned whether she utilizes the implicit path depends on the accuracy of detection.

DISCUSSION

In our discussion, we first make clear the limitations of our evaluation. We understand the outcome of our project is only based on the result of this preliminary study and what authors have learned throughout the project. Then we classify the role of explicit commands in an implicit control loop. Next we present design implications for an implicit control loop.

Limitation of the Study

Due to the fact that several wires were connected to the subject and a laptop, the naturalness of the field evaluation remains questionable. Subjects reported that they sometimes feel embarrassed, but do not change their behaviour. Though we tried hard to make the study as natural as possible, we acknowledge that the power of our study is limited. Also, the small number of subjects in our study brings its validity into questions. We admit that the design for explicit and implicit paths may be very sensitive to the context and individuals differences.

In our project we focus on one specific use case, which is attentional bookmarking (interruption driven) in the context of listening to spoken audio streams when multi-tasking. The value we find while investigating this particular use case is helpful to other use cases as we hope to provide insights on a more general level. Our opinions on whether it can be generalized, and to what extent, are optimistic but conservative.

While we concentrate on exploring explicit and implicit interaction paths in our proposed paradigm (Figure 1), how users receive confirmation or notification of implicit system changes is not our primary task in the project. We have only applied a naively-designed audio feedback rather than designing haptic feedback via a more thoughtful process. Subjects of the preliminary field evaluation indicated that our system notifications (the audio feedback) were poorly designed. This may have impacted the level of users' acceptance on implicit system changes during our study.

Design Implications for Explicit Commands

Explicit commands in an implicit control loop should have the ability to rapidly correct or tune system behaviour. For example, an "undo" function can help the system recover from incorrect changes generated by implicit controls. In a traditional explicit control loop, users should be able to retrieve system information effortlessly. This holds true in an implicit control loop. Information regarding system changes due to implicit paths should be retrievable as well. Explicit paths are considered essential in an implicit control loop.

Explicit paths should offer the user complete functionalities. They not only support what implicit paths cannot achieve, but should also be capable of performing all kinds of commands when the implicit system crashes or lacks user trust. The system should still perform well with only explicit interaction paths. Users should always have full control of the system whenever they want it. Users should be able to choose what explicit commands they want to react to an implicit system change.

Explicit and implicit paths are considered to support each other rather than competing with or replacing each other. Switching between both paths as system inputs should be smooth. Consequently, the characteristics of both paths should be transparent, natural, and similar to each other.

Design Implications for Implicit Control Loop

It requires some effort for users to understand and trust a system operating with an implicit control loop. Similar to how users deal with a new interaction method, they need to learn and adopt the interaction loop formed of implicit controls. On the other hand, a system with an implicit control loop needs time to learn and improve adaptively. One way for the system to achieve better performance is to record and analyze when and why an explicit command is performed. Time is an important factor for both the system and the users.

The accuracy (and appropriateness) when the system predicts and performs a desirable action based on data from implicit inputs is another important factor in the design process. A system should act differently based on how confident it is, and should choose the best time to notify or interact with the user possible. When designing systems with implicit control loops, one should always take the degree of the system's accuracy and user context into consideration.

The advantage of employing implicit control channels is that it requires a lower level of user attention. Designers should keep in mind that any consequences of system changes due to implicit control channels should never be disturbing, so to avoid requiring even more attention than a fully explicit system.

CONCLUSIONS AND FUTURE WORK

In this project we explore the value and design implications of explicit paths in an implicit control loop. We choose attentional bookmarking (interruption driven) for spoken audio streams as our use case. We examine the use case, build and informally evaluate a prototype. We have learned design implications for explicit commands and implicit control loops in this project, as well as the specific way portable players for spoken audio streams should function. This project contributes by giving interesting insights into the design implications for explicit commands and implicit control loops on a general level. The outcome of this project can be the catalyst for further investigation.

The existing prototype can be iteratively modified and commercially produced. Additionally, other design alternatives can be prototyped, tested, and compared with one another. The roles of haptic and audio feedback in this context, including what information to reveal and when to deliver, can lead to another study. A more precise and well structured study can be performed to carefully examine design implications. For example, a longitudinal study can verify and user behaviour changes more thoroughly. It is also possible to apply other use cases and examine if consistency and generality exist.

ACKNOWLEDGEMENT

Thanks to Oliver Schneider and Lily Chen for editing. Special thanks to Sophia Liang for her invaluable support. Lastly, thanks to Tom Hazelton, for providing insightful suggestions on the paper and for editing.

REFERENCES

1. Iqbal, S. T. and Horvitz, E. Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. In *Proc. CHI 2007*, ACM Press (2007), 677-686.
2. Hazelton, T.W. Investigating, Designing, and Validating a Haptic-affect Interaction Loop Using Three Experimental Methods. *Unpublished M.Sc. thesis*, The University of British Columbia (2010).
3. Pan, K.X.J. M., Chang, J., Himmetoglu, L., Moon, A., Hazelton, W. Thomas, MacLean, E. Karon, Croft, A. Elizabeth. Now, where was I? Physiologically Triggered Bookmarking, In *Proc. CHI 2011*, ACM Press (2011).
4. MacLean, K.E. Putting haptics into the ambience. In *Proc. IEEE Transactions on Haptics*, *3*, 2 (2009), 123 – 135.
5. Baumann, M. A., MacLean, K. E., Hazelton, T. W., and McKay, A., Emulating Human Attention-Getting Practices with Wearable Haptics. In *Proc. Haptics Symposium 2010*, IEEE (2010).
6. Apple Magic Mouse. [Online] http://www.apple.com/magicmouse/.
7. Synaptics Gesture Suite for TouchPads. [Online] http://www.synaptics.com/solutions/technology/gestures/touchpad.
8. Irani, P. Blurring boundaries: Mixed input modes to alleviate common limitations with current digital environments. Talk at *SIGCHI PARIS 2011*. (2011).
9. Weiser, M. and Brown, J.S. Designing calm technology. In *Proc. PowerGrid Journal*, 1, 1 (1996), 94-110.
10. Lee, S., Kim, S., Jin, B., Choi, E., Kim, B., Jia, X., Kim, D., Lee, Kun-K. How users manipulate deformable displays as input devices. In *Proc. CHI 2010*, ACM Press (2010), 1647-1656.
11. Saffer, D. Designing Gestural Interfaces. O'reilly (2008).
12. Open Source Gesture Library. http://gestureworks.com/features/open-source-gestures/.
13. 10/GUI. [Online] http://10gui.com/video/.
14. 8Pen. [Online] http://www.the8pen.com/.
15. Mistry, P. and Maes, P. SixthSense: a wearable gestural interface. In *Proc. SIGGRAPH ASIA 2009*, ACM Press (2009).
16. Harrison, C., Tan, D., and Morris, D. Skinput: appropriating the body as an input surface. In *Proc. CHI 2010*, ACM Press (2010), 453-462.
17. Saponas, T. S., Tan, D. S., Morris, Dan, Balakrishnan, Ravin, Turner, Jim, Landay, J. A. Enabling always-avaiable input with muscle-computer interface. In *Proc. UIST 2009*. ACM Press (2009), 167-176.
18. Uikkaniemi, K., Laitinen, T., Turpeinen, M., Saari, T., Kosunen, I., and Ravaja, N. The influence of implicit and explicit biofeedback in first-person shooter games. In *Proc. CHI 2010*, ACM Press (2010), 859-868.
19. Yamaha BodiBeat. [Online] http://www.yamaha.com/bodibeat/.
20. Emotiv. [Online] http://www.emotiv.com/.
21. Nike+iPod. [Online] http://www.apple.com/ipod/nike/.
22. Zheng, Y., Li, Q., Chen, Y., Xie, X., and Ma, W.-Y. Understanding mobility based on GPS data. In *Proc. UbiComp' 2008*. ACM Press (2008), 312-321.
23. Dornbush, S., Fisher, K., McKay, K., Prikhodko, A., and Segall, Z. XPod a human activity and emotion aware mobile music player. In *Proc. International Conference on Mobile Technology, Applications and Systems 2005*. (2005), 1-6.
24. Ju, W., Lee, B., and Klemmer S.R.. Range: Exploring Implicit Interaction through Electronic Whiteboard Design. Poster at *CSCW 2008*. (2008).

25. Ambling BookPlayer. [Online]. http://amblingbookplayer.com/.
26. Fels S., Miller G., and Müller S. Using temporal video annotation as a navigational aid for video browsing. Poster at *UIST 2008*. ACM (2008).
27. Snibbe, S. S., MacLean, K. E., Shaw, R., Roderick, J.B., and Verplank,W. Scheeff, M. Haptic Metaphors for Digital Media. *In Proc. UIST 2001*, ACM Press (2001), 199-208.
28. Bernstein, A. An analysis of the skin conductance OR in samples of American, British, and German schizophrenics. *Biological Psychology 1982*, Elsevier Science B.V. (1982), 14(3-4):155-211.
29. Frith, CD, Allen, HA. The skin conductance OR as an index of attention. *Biological Psychology 1983*, Elsevier Science B.V. (1983), 17(1):27-39.
30. Cacioppo, JT, Tassinary, LG, Berntson, GG. The Electrodermal System. *Handbook of Psychophysiology*, Cambridge University Press, New York, NY. (2007).
31. Healey, J. and Picard, R. StartleCam: a cybernetic wearable camera. Digest of Papers. In *Proc. Second International Symposium on Wearable Computers 1998*, IEEE Comput. Soc (1998), 42-49.
32. Kern, D., Marshall, P., Schmidt, A. Gazemarks: gaze-based visual placeholders to ease attention switching. In *Proc. CHI 2010*. ACM Press (2010), 2093-2102.
33. Hazelton, T. W., Karuei, I., MacLean, K. E., Baumann, M. A., and Pan, M. Presenting a biometrically driven haptic interaction loop. In *Proc. Whole Body Interaction 2010, a SIGCHI 2010 Workshop*, ACM Press (2010).
34. Kim, Jungsoo, He, Jiasheng, Lyons, K., and Starner, T. The Gesture Watch: A Wireless Contact-free Gesture based Wrist Interface. In *Proc. International Symposium on Wearable Computers 2007*, IEEE (2007), 1-8.
35. Apple iPod nano. [Online] http://www.apple.com/ipodnano/.
36. Apple Earphones with Remote and Mic. [Online] http://store.apple.com/us/product/MB770G/B?fnode=MTY1NDA0Ng&mco=MTM3NDk2ODg.
37. Mister Gloves - A Wireless USB Gesture Input System. [Online] http://courses.cit.cornell.edu/ee476/FinalProjects/s2010/ssc88_egl27/index.html.
38. CyberGlove. [Online] http://www.mindflux.com.au/products/vti/cyberglove.html#overview.
39. Jaijongrak, V., Chantasuban, S., Thiemjarus, S. Towards a BSN-based gesture interface for intelligent home applications. In *Proc. ICCAS-SICE 2009*, IEEE (2009), 5613-5617.
40. Pan, M., Baumann, A. M., Hazelton, W. Thomas, MacLean, E. Karon, and Croft, A. Elizabeth. Expressive, Wearable Haptic Displays, Demo in *Haptics Symposium 2010*. IEEE (2010).
41. Affectiva. [Online] http://www.affectiva.com/.
42. Thought Technology. [Online] http://www.thoughttechnology.com.
43. Engineering Acoustics, Inc. C2 Tactor (Data Sheet). [Online]. http://www.eaiinfo.com/EAI/PDF%20Documents/C-2%20tactor.pdf.

Appendix A. Experiment procedure

Gordon Chang, version 3.0

Set-up

- Ask subject to provide at least two interested but not yet listened audio books / podcasts (the researcher will choose one to play), each as least 30 minutes long. Ask subject to bring his earphone. Ask subject to bring valid bus pass.
- Prepare the devices, including the laptop, the prototype, GSR sensor, and video recorder. Make sure the latter three have full battery power.
- Find helpers A, B, C, D, and explain to them what their tasks are.
- Give helper D (the camera man) some time to familiarize with the video recorder.
- Change the selected file name to 1.mp3. Run the program to see if audio output is working.
- Helper D will walk next to but slightly behind the subject to record the subject's interaction on the prototype during the study. He is not visually out of range from the subject.
- The researcher will walk next to bus slightly behind the subject to observe subject's interaction. He is visually out of range from the subject.

Instruction to the Subject

1. **Explain the whole procedure**
   A. Explain the purpose of the study again.
   B. The study will include 4 stages: set-up, practice, the session, post-session interview. (explain each stage)
   C. Tell the subject the commuting route (walk from CS building to Starbucks, from Starbucks (pretending that it's home) to bus stop 59715, take any bus that goes to bus loop, get off at bus loop, walk on East Mall to CS building, take elevator to 5th floor, go inside X508)
   D. During the study, try to be as natural as possible
   E. Inform the subject that the interaction will be recorded by the camera man during the study.
   F. The GSR signal will be recorded and may be used in other pieces of the HALO project.
   G. Instruct the subject to pay close attention to the audio file and aim to understand it fully.
2. **Put devices on the subject**
   A. Put the prototype on the subject's hand (same hand as the subject usually wears his watch).
   B. Put the GSR sensor on the subject's non-dominant hand, and tell the person to not play around with it.
   C. Have the ear phone ready
3. **Explain how the prototype works**
   A. Introduce interaction techniques of fast forward 15 / 60 seconds, rewind 15 / 60 seconds, play, pause, place bookmarks, previous bookmark, next bookmark.
   B. Explain how the Wizard of Oz will work throughout the study.
   C. The subject has full freedom to interact with the device at anytime.
4. **Ways to pause / terminate the study**
   A. Since it is a study in the field, and the user has full freedom of what he can do, the only way to pause or to terminate the study is either the subject or the researcher performs a hand signal ("X" signal, crossing arms in front of the chest)
5. **Go through the heuristic with the subject, modify the list if possible.**
6. **Introduce the camera man**

Experiment Procedure

1. **Introduction and Device Set-up**

   1.1 Set up the devices. Explain to the subject the whole procedure (see above)

   1.2 Test out audio file and the program, to see if the volume is alright

   1.3 Record GSR signal for 15 seconds, see if the signal is valid

   1.4 Run the program and have the subject try out the interaction technique. Make sure he understands every function. Try out few functions. Look at the logger file, and see if it is successfully recorded.

2. **Practice session**

   2.1 Adjust the positions of three – the subject, the researcher, and the camera man

   2.2 Note the timer. Start video recording, start the program, start GSR collection.

   2.3 The subject walks down to Starbucks on Agronomy road (followed by the researcher and the camera man).

   2.4 Stops at Starbucks.

   2.5 Ask the subject if there's any issue or problem. Adjust accordingly.

3. **Evaluation session**

   3.1 Remind the subject from now on, only the terminate arm signal can pause the study. Otherwise, the researcher will pretend that he does not exist.

   3.2 Change the audio file to a new file

   3.3 Throughout the study, the researcher may jot down the global timestamp and write some notes whenever possible.

   3.4 The subject follows the designed route until reaches the destination.

4. **Interview session**

   4.1 Stop the program, stop the GSR collection, stop video recording.

   4.2 Thanks the camera man. Take off all the devices except the prototype from the subject.

   4.3 The researcher opens his laptop, start interviewing the subject.

   4.4 Take off the prototype and thank the subject for participating.

Appendix B. Heuristic

Gordon Chang, version 3.0

The interruption is based on following observation from the researcher. The researcher will judge whether the event is a valid interruption.

- **Abnormal or large body movements**
  - Including the body, fingers, sudden change of walking speed (starts walking, stops, etc), etc
  - Excluding when the subject is controlling the device
- **Sudden change of facial expression**
  - Whenever the subject frowns, feels surprised, looks happy, etc
- **Orienting response**
  - Definition: immediate response to a change in its environment, where the subject pays attention to it even before the subject identifies it
  - Usually the person may turn his/her head
  - Typical events are a sudden sound or noise in the environment, or moving object(s) in subject's sight
  - Examples: people passing and chatting, a cute girl passing by, construction noise, bicycles/cars passing by, someone starts talking to the subject, etc
  - The events are valid only when they catch the subject's attention

Sample events during the experiment

The researcher will look closely on following events to see if it counts as an interruption. Examples:

- Bus coming (when the subject notices it)
- The subject steps on the bus
- The subject shows bus ID
- The subject sits on a seat
- The subject notices the stop and gets up from the seat
- Bus door opens
- The subject starts walking from the bus stop
- The subject takes out object(s) from his pocket / backpack
- Earphone unplugged
- Lab door or entrance door of the building opens (either by someone or by the subject himself)
- etc

Appendix C. Interruption set-up

Gordon Chang, version 3.0

Artificial interruption events

1. *Phone call from a friend*

   Helper A, a friend of the subject, makes a phone call to the subject, talks with the subject for about 30 seconds to a minute, after the subject gets off from the bus.

2. *Stranger asking for directions*

   Helper B, a tourist, who walks on East Mall, approaches the subjects and asks where the SUB is, in detail. The conversation is planned to last 20 to 40 seconds.

3. *Bumping into a lab-mate / friend*

   Helper C, a lab-mate who's also going to the lab, bumps into the subject, has approximately one minute chat with the subject, and takes the same elevator with the subject.

   Helper C, a lab-mate who's going to the lab/office, bumps into the subject, has approximately one minute chat with the subject, and say good bye to the subject.

   Helper C, a friend, runs across the subject near the department building, has a small chat (approximately one minute) with the subject, and leaves.

Designed natural interruption events

1. *Subject gets on the bus and seeks a seat*
2. *Subject walks by a construction site in progress*

Other artificial events

1. *Fake bookmarks*

Involved people for the study: a total of 6

<u>The researcher</u>

<u>The subject</u>

<u>Helper A</u>: Someone who knows the subject and has subject's phone number

<u>Helper B</u>: Must not be recognized by the subject, them should not know each other

<u>Helper C</u>: A friend who knows the subject and works in the same building (or not)

<u>Helper D</u>: The camera man, a person to video record the subject's interaction on the prototype during the study

The helpers are told in advance that they need to pretend the video recorder and the researcher do not exist.