Controlling Smart TVs using Touch Gestures on Mobile Devices

Jiaxu Sun, Yongchao Li, Linzhang Wang, Xuandong Li
State Key Laboratory for Novel Software Technology
Nanjing University
{sunjiaxu, lyc}@seg.nju.edu.cn, {lzwang, lxd}@nju.edu.cn

Xiaoxiao Ma, Jing Xu, Guanling Chen
Department of Computer Science
University of Massachusetts Lowell
{xma, jxu, glchen}@cs.uml.edu

***Abstract*—The global TV market has experienced significant changes in its market landscape, and the Smart TV market is growing fast. However, a traditional TV remote is not suitable for interacting with many Smart TV applications; this is particularly true when typing text is necessary, such as searching, browsing, and using social applications. On the other hand, increasingly-pervasive mobile devices, scuh as smartphones, are emerging as an interesting platform for TV interactions given their intuitive input and output modalities.**

In this paper, we present and evaluate a new design of Second-Screen Smart TV control app, *S3TV*, for personal mobile devices. We focus on how to reduce visual attention shifting by leveraging touch gesture controls, tailored for peronsal mobile devices. We conducted comprehensive user studies, subjective usability analysis, objective interaction analysis, and comparative studies against a soft button based control app. The subjective usability analysis showed the prototype achieved a System Usability Score (SUS) of 80.21, well above the average score of 70.12. In addition, the objective interaction analysis confirmed high task success rates and correct action sequences, and touch gesture controls were preferred over hard system keys. Compared to de facto soft button-based control approach, *S3TV* reduced visual attention shifting and operation delay with improved user experience. Overall the participants reported that they felt satisfied and would like to continue to use *S3TV*.

***Keywords*-Smart TV; Second-Screen Apps; Interaction Design; Usability Testing;**

I. Introduction

Smart TV has become a fast-growing market due to the consumer's demand for multi-functional services, the provider's opportunity for new revenues, and the bandwidth increase of broadband Internet [1]. Second-screen, sometimes also referred to as "companion device", is a term that refers to an additional electronic device (particularly personal mobile devices such as tablets and smartphones) that allows the viewers to interact with the content they are consuming [2]. On the other hand, the interaction between a viewer and a Smart TV with a traditional remote control is problematic, especially for complex operations such as searching with keywords [3]. It has led to the proposal of utilizing second-screen devices, not solely for consuming extra content, but also as TV controlling devices to replace traditional remote controls [4]. The benefits of using modern mobile devices as TV remote controls include: better information visualization (information shown on second-screen devices is closer to users for easier reading), multi-modal inputs (touch-screen gestures, virtual keyboards, and voice recognition), and multitasking with mobile apps without interrupting TV experience.

Existing second-screen mobile apps to *control* a Smart TV mostly use "soft buttons," such as the "Able Remote" [4]. While recent studies suggest that touch sensitive control apps can outperform traditional remote controls [5], a major limitation is that users have to shift their attention from the TV display to the remote control, which may cause an unpleasant delay in the operation [3]. Comparing to traditional remote controls, many errors were triggered because of the unintentional activation of a command (e.g. soft buttons on the touch-sensitive small screen). So far little work has been done to design and evaluate a touch-screen based solution that address this problem, which is considered by many viewers to be a significant drawback of second-screen apps [6].

The main purpose of this paper is to present and evaluate the novel design of *S3TV*, a Second-Screen app for mobile devices that leverages three design principles: 1) separate "content" and "control" interfaces so the user can maintain a clean mental model for using the mobile device for content consumption and as a remote control, 2) promote full-screen touch gestures for video playback controls, and 3) design cross-screen UI elements, such as a draggable video progress bar, by considering the user's operational context. The combination of these designs helps users reduce visual attention shifting for enhanced user experience, thus better positioning mobile devices as more intuitive and effective TV controllers.

The novelty of S3TV lies in its unique design of leveraging full-screen touch gestures for multi-screen interactions and we studied design alternatives such as using overloaded system keys. To the best of our knowledge, this is the first work to specifically address the visual attention shifting issue for second screen interactions with comprehensive user studies.. The overall user experience survey results showed that *S3TV* was well received and considered favorable to existing methods such as traditional remote controls or de facto soft button based control mobile apps. The touch gesture-focused controls that take advantage of touch-screens were well liked. The detailed subjective usability evaluation showed that the System Usability Scale (SUS) of *S3TV* reached 80.21, well above the average score of 70.12, suggesting that *S3TV* is effective and satisfactory. In addition to the qualitative user study, we also conducted a quantitive analysis of all UI interaction events, which confirmed that most participants correctly executed task sequences. The analytic results also suggest that participants prefer touch-screen gestures to overloading system keys for design consistency reasons. Finally, we compared *S3TV* with a soft-button based remote control app. By analyzing the post-study surveys and reviewing the video recordings of the participants during the study, we confirmed that *S3TV* was successful in reducing the overhead of visual attention shifting between multiple devices.

The contributions of this paper include:

- An innovative design of a second-screen app that provides a full-screen touch gesture controls for Smart TV operations to reduce visual attention shifting;
- A reference implementation of the new design with an Android-based *S3TV* prototype for off-the-shelf devices, and we explored and compared the design alternatives leveraging the system keys on these devices;
- Comprehensive user studies for both subjective evaluation using System Usability Scores (SUS) [7] and objective interaction analysis using an Android usability toolkit [8], with reported experience and lessons learned to influence future design of such apps.

The rest of paper is organized as follows. In Section 2, we discuss related works on second-screen research. We then present the design and implementation of *S3TV* in Section 3, and describe our user study experiment setup in Section 4. We present overall user performance and subjective usability evaluation, objective analysis of interaction events, and comparative study with a soft-button control app in Section 5. We conclude and discuss plans for future work in Section 6.

II. Related Work

It has been shown that interaction with smart TVs can impose usability challenges [3], particularly for the elderly who may have problems when navigating through the application interface on the TV [9]. Several researchers have proposed to use touch sensitive devices to address this challenge. Cruickshank et al. designed a second-screen app on a PDA, eliminating the need for the TV to use on-screen graphics [3]. The experimental results show that though the system fell short on providing tactile feedback, users still showed a strong preference for it over a conventional remote control. Though the results have also confirmed another major limitation with using mobile touch-screen devices as remote controls: users have to shift their attention from the TV display to the remote control, which may cause an unpleasant delay in the operation. Comparing to traditional remote controls, Pirker et al. found that most errors caused by touch sensitive remote controls were related to precision problems with the touch interaction, overlooking of navigation items, problems with speed or accuracy of the touch interaction, the unintentional activation of a command, and general usage problems like slips or misactivation because of the unaccustomed interaction technique [10]. Bernhaupt et al. provides a survey of interaction techniques related to mobile phones: voice/speech, mid-air gesture, and touch gesture [11]. While voice/speech interaction may still need some time for technology to mature, mid-air gesture based on device movement is less efficient though it does not remove user's visual attention away from the TV. On the other hand, touch-screens haves become an acceptable means of interaction to control TVs, though the drawback has also been acknowledged that users have to shift their visual attention [12].

There are several existing mobile apps acting as TV remotes. For instance, "AirPlay" enables iOS devices to send content to an HDTV through Apple TV, and turns those iOS devices into second-screen controllers [13]. However, once the user exits the interface to use other apps, the connection to the TV breaks and the content on the TV disappears. While "AirPlay" is quite popular, the viewer can easily get fatigued from frequently switching their visual attention between the TV screen (high position) and the mobile screen (low position) [6]. Many Google TVs come with trackpad-equipped remote controls that allow for one-thumb operations, though text entry can be a painful experience. "Able Remote" is a popular remote control app for Google TV [4]. However, the complex interface is not intuitive [6] and it requires a lot of visual attention in order for users to find the position of a desired soft button on the small smartphone screen.

These results from the relevant literature and existing apps suggest: 1) second-screen mobile apps for TV interaction are desirable; 2) touch-screen gestures are preferred means of interaction; 3) second-screen apps for TV interaction lack tactile feedback and introduce operation delays; and 4) existing touch-based apps based on soft buttons are not optimized for user experience. Thus we are motivated to design *S3TV* to take advantages of touch-screen gestures while reducing the visual attention shifting through full-screen gestural interface and on-TV operational feedback.

III. S3TV Design and Implementation

In this section, we describe the design of the proposed Second-Screen Smart TV application, *S3TV*, which runs on personal mobile devices, such as touch-screen smartphones and tablets.

A. Design Principles to Improve Multi-Device Experience

We followed three design principles motivated by existing research and our preliminary studies to optimize the user's experience.

Principle 1: The *S3TV* application should have separate "content" and "control" modes; so the user can maintain a clear mental model of the smartphone acting as a content consumption device (e.g. description of the show or other related content) or a remote control (e.g. fast-forward/rewind, pause/resume, etc.). "Able Remote" mixes content buttons and control buttons on the same screen [4], and thus users must shift attention from the device (to click the right buttons) to the TV (to see the feedback of the action, such as fast-forwarding), which results in back-and-forth eye focus and unintentional activation of a command [10]. On the other hand, WeBet presents a dedicated control interface for users to place bets using gestures while keeping their attention on the TV [14]. Their results support our design choice of the separation of content and control modes to allow for a better operational experience.

Principle 2: The control interface of the *S3TV* application should promote touch gesture-based control actions to reduce visual attention shifting. Nowadays, consumers have already gotten used to touch-screen gestures such as tapping, swiping, and dragging on modern smartphones and tablets. These gestures do not consume significant visual attention, so the user can focus instead on the TV. Existing literatures show that touch-screen gestures work more intuitively and efficiently than button-based interfaces [15], [16], [5]. However, Nielsen and Budiu caution that gestures should follow conventions, and that overuse of gestures can challenge usability [17].

Principle 3: The cross-screen UI element must be designed to reflect the user's operational context, as the user switches their visual attention between the TV and their mobile device. For instance, the gesture control of video playback should have feedback on the TV, and the gesture should be recognized anywhere on the mobile screen without needing to look at the mobile device. The video play button should have "Play on TV" or a similar label so the user understands where the video will be shown.

These design principles help users to achieve a clear mental model when using *S3TV*. Figure 1 shows a user interacting with *S3TV* in content mode, which requires her visual attention. This is typical for users to consume additional content of the TV show, to search for a new show, or to share the show with her social networks. Figure 2 shows a user operating *S3TV* under control mode for video playback operations without lowering her head to read the phone screen, very much like how she would use a conventional remote control.

Fig. 1. Content mode requires visual attention.

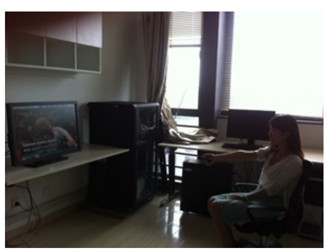

Fig. 2. Control mode with eyes-free operations.

B. S3TV Implementation and Video-Playback Gestures

For prototyping purposes, we chose the Android platform as it has rapidly become the fastest-growing mobile OS. To evaluate the proposed interaction and control designs, *S3TV* implementation is currently simplified to focus on the separation of content and control modes, and the touch gesture-based video playback operations. We paid less attention to designing a rich interface for channel/show information display and selection, and other convenience type features. Instead we focused on control interface to evaluate the effectiveness of the full-screen touch gestures.

When the *S3TV* app starts on the mobile device, it searches the local Wi-Fi network to discover and connect to the Smart TV when available. The app then enters content mode and downloads a video list from the TV (video contents are stored on the TV). The user can then search or browse the video information on their mobile device, taking advantage of the virtual keyboard or voice-based text entry on the mobile device. When clicking the "Play on TV" button, the *S3TV* app will enter the control mode, in which the interface has a simple background picture showing a list of supported gesture operations. We expect that users will become familiar with the gestures, and so they will not need to read the instructions frequently as long as the gestures are intuitive and easy to remember.

We describe the supported video playback gesture-based operations of control mode as follows. Note we chose to only implement a minimal design as previous study showed that users only need a small set of controls for their smart TV experience [16].

Double tap for pause and play: Double tapping is defined as two consecutive touches on the screen which will switch media between pause and play. In our earlier design, a single tap was used to pause and play. The user testing, however, showed that an accidental touch frequently triggered this action, particularly when the user was watching the TV. Thus, we required double tapping for this action to avoid misactivation.

Drag to a particular video position: Existing video applications often have a slider bar enabling the user to drag the video to a particular position to watch. If this was directly implemented on the mobile device, however, it would require a head-down operation for the user to precisely drag the slider button along the progress bar. Instead, we allow the user to drag along the whole screen of the mobile device without any visible slider bar. The user can continue watching the TV and simply slide their finger anywhere on the mobile device. The video on the TV will then move forward/backward, following the finger movement, and the TV will display a progress bar sliding synchronously as a visual aide.

Swipe left or right to fast-forward or rewind: When the user continues swiping left or right, the rewind or fast-forward speed will increase (2X, 3X, 4X, etc.) The TV displays nX (n=2,3,4..) as a visual feedback, so the user does not need to look at the mobile device. This process can be stopped and the video will resume to play with a double tap. Note that swiping is a one-time flicking gesture, and it is different than continuous dragging on the screen.

Swipe up or down for volume adjustment: The user can swipe up and down to adjust TV volume, similar to the previous swiping operation noted for fast-forward/rewind. In addition, as modern mobile devices have hard keys for adjusting the volume of the device, we also overloaded these volume keys so that they will adjust volume of the TV when running the *S3TV* application. We are interested in finding out which method the users prefer.

V gesture for taking screenshot: There are other operations that can be applied on the video being played, such as the commonly requested feature of taking screenshots of the TV show. *S3TV* allows the user to draw a "V" gesture on the mobile device's touch-screen for this feature, which will work regardless of whether the video is paused or not. This is a special gesture that the user is often unfamiliar with and thus has to learn it. We are interested in seeing how a less intuitive gesture is accepted in such an environment. We have also implemented this function when the user presses the hard CAMERA key, when available on the Android devices.

Long press for menu of additional operations: We cannot expect to design new gestures for all the possible operations, such as recording the current video, adding to favorite channels/shows, or sharing the video with social networks. The user will only be able to learn and recall a few special gestures, such as the "V" gesture to take screenshots. Thus, for operations infrequently used, we implemented an action menu that pops out when the user long-presses the touch-screen. Note that this will be a head-down operation, as the user has to pick a menu item. In this case, it is, however, acceptable for the user to remove visual attention from TV as they may intend to conduct a complex task, such as to write a comment and share to a social network.

Shake the device to change video: To exercise different kinds of gestures, we implemented "shaking" the device as a method to change the current video to the next one on the list. The shaking gesture is based on accelerometer readings and is already popular in mobile applications and games. Alternatively, the user can exit *S3TV* control mode and select a different video while in the content mode.

Back key to exit control screen: Android devices have a hard BACK key that is used to exit the control screen and place the device into content mode. The video will continue to play on the TV while the user can now search/browse for new videos, read additional information about the show, or even start using other applications on the device. The content screen of *S3TV* has a "Control" button that allows the user to get back into control mode to operate the video playback.

Display incoming calls on the TV: *S3TV* is designed to run on the user's personal mobile device, and an incoming call has a high priority that can interrupt other applications. Consider that the user may not hold the phone the entire time when watching the TV, and in the possibly loud environment, incoming calls may not be noticed. *S3TV* can display the incoming call information on the TV as a convenience when the mobile device and the TV are connected. While this feature was liked by our test participants, it is not the focus of this paper and we will formally evaluate its effectiveness as future work.

Note that in *S3TV*, some hard system keys are used as video playback operations. The overloaded usage of these hard keys, however, is designed to be consistent with their original intended usage to reduce any learning curve. For instance, the volume adjustment keys are used to turn up/down the TV volume with the *S3TV* app. We also provided corresponding gestures (e.g. swipe up/down for volume adjustment) in case the users find the overloaded hard key functions confusing.

IV. User Study

We recruited 12 participants, which is recommended by the System Usability Scale evaluation [7]. All participants were graduates or undergraduate students, and three of which were Computer Science students. Before the experiment, we collected basic demographic data from the participants on their major, cellphone platform, familiarity with the Android platform, and their main uses of the mobile phone. The surveys showed that 11 participants had used both feature phones and smartphones, while 1 student had used only feature phones. All of the 11 smartphone participants were familiar with the Android. In addition, 5 participants were familiar with other mobile systems, such as iOS and Symbian. All 12 participants had some experience with touch-screen controls, and 11 participants typically used a mobile phone for more than 1 hour each day. Their usage of mobile phones included web browsing, social networking and reading e-books. Additionally, 10 participants had some knowledge of Smart TVs, and 7 of them had used a Smart TV before. Some participants complained of the cumbersome interaction using their conventional remote controls before their testing of the *S3TV* application.

Each participant was asked to perform the tasks listed in Table I in the same order. Every task was designed to evaluate a certain aspect of the second-screen interactive model. Task T1 was designed to allow participants to become familiar with the system. In T2, we wanted to see how effectively participants could switch between *S3TV* and other apps. In T7, our goal was to study the benefits of combining the advantages of modern touch-screen phones and the Smart TV experience. Other tasks were mainly designed to evaluate the effectiveness of the touch gesture controls. Between different tasks, there was a transition between content and control modes. For example, the *S3TV* enters into control mode in T1, and then switches into content mode from control mode after T4.

Before conducting these tasks, the participants were given 5 minutes to play with the *S3TV*. They were allowed to explore and ask any questions regarding the app. After the learning phase, the participants started to perform the tasks. During the user testing, they attempted to complete each task without assistance. Participants were asked to tell us when they started and finished every task. They were also asked to give the phone to one of the evaluators after completing every task and the evaluator would then assign the next task.

For comparative studies, we also implemented a Remote Control App (RCA).The only difference between the *RCA* and *S3TV* is the control interface: *S3TV* uses full-screen gestures, while *RCA* uses soft buttons to simulate interfaces like "Able Remote" and "AirPlay."Once the participants completed all tasks with *S3TV*, we asked participants to do the same tasks with *RCA*. Finally, every participant was asked to fill out both the SUS and the post-study surveys.

TABLE I
THE LIST OF TASKS FOR THE S3TV EVALUATION USER STUDY.

No.	Task Description	Task Purpose
T1	Select a video and start watching.	get familiar with *S3TV*
T2	When watching TV, pause the video and make a phone call; then return to *S3TV* and continue to watch the video.	switch between *S3TV* and other apps; transition between content and control modes
T3	Select a video and start watching, drag the progress bar to about three-quarters; then continue to watch, drag the progress bar to about one quarter and continue to watch.	test horizontal scrolling gesture
T4	Turn up the volume when watching TV; then continue to watch, and turn down the volume.	test flick up/down gesture and volume key alternative
T5	Select a video and start watching; then select another video and start watching.	transition between content and control modes and test "shaking" gesture
T6	When watching TV, take a screenshot from the current screen and then continue watching.	test a customized gesture that requires learning
T7	Select a video and start watching; then share video information to social networks.	test long-press gesture and menu selection

V. Evaluation Results

We conducted extensive evaluation to answer two basic research questions (RQ):

RQ1: How well does S3TV work as a companion app that allows users to complete typical content and control tasks?

RQ2: How well does S3TV design reduce visual attention shifting compared to a more traditional approach (e.g. RCA)?

For RQ1, we describe the results from three measurements: task completion rate, SUS scores, and interaction events analysis (Section V-A-V-C). For RQ2, we describe the comparison results of *S3TV* against a soft-button based control interface (e.g. RCA) with an analysis from the post-testing questionnaire (Section V-D).

A. Task Performance Analysis

During usability testing, 12 participants performed 7 tasks in the same order. Evaluators recorded the execution time of each task. Only Task T4 had the pass rate of 83%, while the other 6 tasks had completion rates of 100%. This suggests that it was not difficult to perform these basic operations. In T4, all participants attempted swiping up or down to adjust the volume after performing the previous task, which was a horizontal drag to a particular position. The main reason that T4 was more difficult is that before T4, all important functions could be performed with gestures, rather than using system keys. Two participants failed T4 because if the user's finger moves too slowly on the screen, the app could not recognize this gesture, and so the volume was not adjusted.

There were some problems caused by the user being unfamiliar with an Android device. For example, 1 participant who did not know how to return to *S3TV* spent a long time on Task T2. The most time-consuming task was T5. Initially, the task was designed to simulate channel switching on TV, where the gesture of shaking device would play the next video. Alternatively, the user could switch from control mode to content mode to select a new video. During testing, many participants briefly hesitated on the control interface, and then returned to the content interface cautiously. This showed that the shaking device gesture, which is a bit unusual (although engaging), requires more effort to learn and recall in practice.

B. Subjective SUS Scores

The System Usability Scale (SUS) was proposed by John Brooke [7], which is technology independent and has already been tested on hardware, consumer software, websites, cellphones, interactive voice response systems (IVRs), and even the yellow-pages. It has become an industry standard with references in over 600 publications [18]. Some minor modification of SUS was made in practice, for example, Bangor et al. found some confusion with the word "cumbersome," and replaced it with "awkward." They also replaced the word "system" with "product" in SUS [19]. According to these researchers, SUS may provide developers with powerful information about product usability. Similarly, the SUS has proven itself a valuable and robust tool in helping to evaluate the quality of user interfaces [19].

In our study, we used modified SUS statements proposed by Aaron Bangor, who developed a reliable, low-cost usability scale to evaluate the system's usability. The SUS is a ten item questionnaire with five response options:

(1) I think that I would like to use this product frequently;
(2) I found the product unnecessarily complex;
(3) I thought the product was easy to use;
(4) I think that I would need the support of a technical person to be able to use this product;
(5) I found the various functions in this product were well integrated;
(6) I thought there was too much inconsistency in this product;
(7) I would imagine that most people would learn to use this product very quickly;
(8) I found the product very awkward to use;
(9) I felt very confident using the product;
(10) I needed to learn a lot of things before I could get going with the product.

Aaron Bangor and Philip T. Kortum have presented nearly 10 years' worth of SUS data collected on numerous products in all phases of the development lifecycle. They began collecting data with the SUS in 1996, and 2,324 surveys have been completed over the course of 206 studies. They conclude that the average SUS score is 70.14, with a 99.9% confidence interval ranging from 68.7 to 71.5. This means that products that are at least passable have SUS scores above 70, while better products are scoring in the high 70s to upper 80s. As such a SUS score of 90 to 100 is equivalent to an A, 80 to 89 a B, and so on. They show that the results are accurate over 90% of the time when there are at least 12 participants. In Table II, our mean SUS score of 12 participants is 80.21, which is far above average. Despite being only a prototype system, *S3TV* shows highly satisfactory usability and demonstrates the effectiveness of the proposed designs.

The SUS average scores of Statement 6 and Statement 8 were less than 3, meaning the confusion level here was higher than others, as none of the participants was familiar with *S3TV*. Statement 1 received the highest score, reflecting that the participants were more satisfied with *S3TV* when compared to existing remote controls. The full screen touch gestures of *S3TV* were praised by the participants in their responses due to reduced shifting of visual attention. The scores for Statements 2 and 10 were 3.0, suggesting that the participants did not prefer overly complex second-screen applications that have both control and content features. Thus the designers of such apps must trade-off the simplicity and what features to include on the companion devices[16].

C. Objective Interaction Analysis

To further validate the effectiveness of our proposed design principles and to identify the usability problems of this app, we used a UI event capturing and analysis toolkit during the user study [8]. The toolkit was embedded into the *S3TV* app and was run in the background to automatically collect user interaction events. These logged user interactions are divided into sequences of user study tasks and transferred to the backend server for quantitative analysis.

We generated 84 sequences (7 tasks for each of 12 participants). We lost 1 sequence due to a manual mistake, and the following analysis is thus based on the remaining 83 sequences. Though we were interested in finding general usability issues of the *S3TV* app, we focused on validating whether the gesture design was efficient and usable. Hence, to distinguish the two issues that we planned to address, we separated our quantitative measurement into two parts: with one involving examining the app's workflow, and the other evaluating the design of gesture control. It is easy to separate the two parts as all gesture control related events are in one activity (control mode). As such, we just need to use the events in this activity for gesture control evaluation, and use the rest of these events for application workflow inspection.

1) Application Workflow: The seven tasks asked users to control video using different gestures, but they shared the same workflow, requiring users to select a video and then watch it. This workflow consists of 4 steps to complete, thus we granulated our analysis to each step so that we can find out at which step users are more prone to make mistakes or get confused. This helps to further locate the potential usability issues.

As we observed from the sequence data, users seldom experienced confusion during the user study, which means that the layout and logic of this application are easy and clear for users to understand. The only observed problem reported by the analytic toolkit was the first step. By taking a closer look, we found that several users pressed the button "Start Smart TV" twice (they only need to press once and then be directed to the next screen). This was likely caused by network delay between the mobile device and Smart TV, so that users became impatient and pressed the button the second time since they were not directed to the next screen after the first button press. This problem can be easily alleviated by providing a visual feedback about network delay, such as a waiting hourglass.

2) Gesture Control: The 7 tasks asked users to use a total of 5 gestures to control the video, which include double tap, scrolling horizontally, swiping vertically, drawing a V pattern, and long pressing. By examining the number of trial actions and the number of successful actions for each gesture, we made the following observations.

(i) Users are used to the double tap, scroll horizontally, and long press gestures, as almost all users completed these gestures with only one action.

(ii) Users experienced difficulty in performing the vertical swipe gesture. There were 2 participants failed to complete this task, and among the 10 users who completed this task, 7 used swipe gestures (the other 3 used the volume keys) and their average number of trial actions is 12.1 times (ideally we would need only two actions: one swipe up, and the other swipe down). The reason is that if a user's finger moves too slowly on the screen, *S3TV* would not recognize this gesture and the volume will not adjust.

(iii) Drawing a V pattern was not hard for users to accept and accomplish. Though 4 users tried more than once to make this gesture (those trial gestures were recognized as swipe gestures), and 7 users succeeded in only one action (and 1 user completed this task with the CAMERA key).

(iv) It seems that users prefer using touch gestures to using the Android's built-in keys when they want to control the video. Volume up and down keys were available as an alternative to making the vertical swipe gesture, and the CAMERA key was included as an alternative to drawing the V pattern. We observed that 6 users tried to adjust the volume with the swipe gesture, while only 3 users used volume keys. And 1 user tried both methods. In order to take a screenshot, 10 users drew the V pattern on the screen, while only 1 user pressed the CAMERA key. And 1 user again tried both methods. Intuitively using the device's built-in hard keys is faster and more convenient than performing gestures on the screen, but the reason that more users chose touch gestures may be due to the following: performing gestures was the only method used in other tasks, users assumed this to be the default method. On the other hand, some users were not familiar with the Android phone, or more particularly, the testing device, so they deem performing gestures as the simpler and more general method. Thus it is desirable to provide consistent gesture-control for a group of operations (video play back) to reduce confusion and learning overhead.

TABLE II
SUS SCORE

No.	1	2	3	4	5	6	7	8	9	10	Score
01	4	3	3	4	3	3	2	3	4	3	80
02	3	2	3	4	4	3	4	3	4	3	82.5
03	3	2	4	3	3	3	3	2	2	3	70
04	3	3	3	3	4	4	4	2	4	3	82.5
05	3	3	3	3	4	3	2	3	3	3	75
06	4	4	3	4	4	3	4	2	3	2	80
07	4	3	3	4	4	2	4	4	4	3	87.5
08	3	4	4	2	3	3	3	3	4	2	77.5
09	4	2	3	4	3	3	3	3	3	2	75
10	4	3	3	3	4	3	2	3	3	4	80
11	4	4	2	5	3	2	4	3	3	4	85
12	4	3	4	4	3	2	4	3	4	4	87.5
Avg.	3.58	3.00	3.17	3.50	3.50	2.83	3.25	2.83	3.42	3.00	80.21
SD.	0.53	0.56	0.45	0.71	0.55	0.45	0.82	0.45	0.64	0.56	

D. Comparison Results with Soft-Button Designs

S3TV has the benefit of reducing visual attention shifting using a dedicated control interface that captures full-screen touch gestures. The feedback of these gestures is shown directly on the TV. Thus, users can continue watching TV while performing gesture controls without looking at the mobile device. This design contrasts with traditional remote controls and emerging second-screen mobile apps, such as "Able Remote" and "AirPlay". For comparative evaluation, we implemented a separate Remote Control App (RCA) whose only difference from S3TV is that its control interface uses soft buttons.

We asked our participants to complete the same tasks using both apps. Participants filled out a post-testing survey to reflect their overall experience using *S3TV*, focusing on evaluating the effectiveness of our *S3TV* design goals of maintaining the user's mental model and promoting gesture-based control, as compared to the soft button based *RCA*. The survey had 10 questions with 5-point Likert scale response options (1 being strongly disagree and 5 being strongly agree):

(1) I would like to use a Second Screen Smart TV app as the remote control.
(2) I prefer this user experience rather than that of traditional remote control.
(3) With *S3TV*, I need not to look down the remote control during its operation.
(4) *S3TV* could reduce the amount of visual attention switching between screens.
(5) With *RCA*, I need not to look down the remote control during the operation.
(6) *RCA* could reduce the amount of visual attention switching between screens.
(7) Compared with *RCA*, *S3TV* may preferably reduce the amount of visual attention switching.
(8) Compared with *RCA*, *S3TV* is the desired design.
(9) The feedback on the TV helps strengthen my viewing experience.
(10) The feedback on the TV helps reduce the amount of visual attention switching between screens.

Q1 (Mean=3.2, SD=1.7): results show that participants would like to use a Second-Screen Smart TV app as they already own, and are familiar with, mobile apps.

Q2 (Mean=3.2, SD=0.8): the majority of participants stated that they would prefer *S3TV* over the traditional remote control. One participant was concerned that several people in the same room could install the *S3TV* app, causing the confusion of who owns the control. This is an issue to be addressed in the future design.

Q3 (Mean=3.8, SD=0.4) and **Q4** (Mean=4.4, SD=0.6): most participants realized that touch gesture-based control could reduce shifting visual attention between screens. Participants particularly liked unconstrained dragging on the mobile screen with feedback showing on the TV. On the other hand, several people still read the mobile screen while drawing a "V" gesture to take a screenshot, as this was not a familiar gesture and it required practice.

Q5 (Mean=2.6, SD=0.9) and **Q6** (Mean=2.4, SD=1.1): scores were lower than that of the identical *S3TV* questions. Participants thought that using soft-buttons consumed a lot of visual attention in order for them to find the position of a desired widget on the small smartphone screen. They were particularly concerned with pressing the wrong buttons.

Q7 (Mean=4.4, SD=0.4) and **Q8** (Mean=4.2, SD=0.8): both of which directly reflect the effectiveness of the touch gesture operations, and all participants preferred *S3TV* over *RCA*. Some of the participants also thought that *RCA* was acceptable, though it does increase operation delays and cause potential misactiviation of commands.

Q9 (Mean=4.4, SD=0.6): participants responded that feedback on the TV could enhance user experience. This is particularly important for touch-screen controls as they do not provide tactile feedback.

Q10 (Mean=3.8, SD=0.4): participant response indicated that they also thought that feedback on the TV contributed to reducing visual attention switching.

VI. Conclusions and Future Work

We conducted comprehensive user studies using a reference implementation of the proposed design, the *S3TV* app, on Android devices. The subjective usability evaluation results show that despite it being an unpolished prototype, the *S3TV* app achieved an overall System Usability Score (SUS) of 80.21, well above the average score of 70.12. We found that users who are familiar with the Android system can easily learn how to use the gesture controls of *S3TV*. In addition, the objective interaction analysis indicated high task success rates and correct action sequences. This suggests that users had little to no confusion of the app design and maintained the right mental models. The results also showed that touchscreen gesture controls were preferred over hard system keys, such as for volume adjustment and taking screenshots. Compared to de facto soft button-based control apps, *S3TV* reduces visual attention shifting and operation delay, thus improving user experience. Overall the participants reported that they felt satisfied with *S3TV* and would like to use it frequently.

Smart TV is still a young platform and there are many exciting research opportunities. We are interested in continuing studying second-screen apps, in particular how to design cross-screen interactions that will reduce visual attention shifting. In addition, we are interested in learning how to leverage more context and devices, such as situations when there are multiple users watching the TV who are, interacting with their own personal mobile devices, to provide a better user experience.

ACKNOWLEDGMENT

This material is based upon work supported in part by the National Natural Science Foundation of China (No. 91318301, 61321491, and 61170066) and the National 863 High-Tech Program of China (No. 2012AA011205). This work is also supported in part by the National Science Foundation of the U.S. under Grant No. 1016823. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies.

REFERENCES

[1] C. Frei, "New market research report – global smart tv market to grow 20.8% from 2012-2017," http://tinyurl.com/ayulukd, 2013.

[2] M. Lochrie and P. Coulton, "Sharing the viewing experience through second screens," in *Proceedings of the 10th European Conference on Interactive Tv and Video*, ser. EuroiTV '12. New York, NY, USA: ACM, 2012, pp. 199–202.

[3] L. Cruickshank, E. Tsekleves, R. Whitham, A. Hill, and K. Kondo, "Making interactive tv easier to use: Interface design for a second screen approach," *The Design Journal*, vol. 10, no. 3, pp. 41–53, 2007.

[4] AbleRemote, "Able remote app," http://ableremote.com/, 2014.

[5] J. Bobeth, J. Schrammel, S. Deutsch, M. Klein, M. Drobics, C. Hochleitner, and M. Tscheligi, "Tablet, gestures, remote control?: Influence of age on performance and user experience with itv applications," in *Proceedings of the 2014 ACM International Conference on Interactive Experiences for TV and Online Video*, ser. TVX '14, 2014, pp. 139–146.

[6] M. E. Holmes, S. Josephson, and R. E. Carney, "Visual attention to television programs with a second-screen application," in *Proceedings of the Symposium on Eye Tracking Research and Applications*, ser. ETRA '12. New York, NY, USA: ACM, 2012, pp. 397–400.

[7] J. Brooke, "Sus-a quick and dirty usability scale," *Usability evaluation in industry*, vol. 189, p. 194, 1996.

[8] X. Ma, B. Yan, G. Chen, C. Zhang, K. Huang, and J. Drury, "A toolkit for usability testing of mobile applications," in *Mobile Computing, Applications, and Services*, ser. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, J. Zhang, J. Wilkiewicz, and A. Nahapetian, Eds. Springer Berlin Heidelberg, 2012, vol. 95, pp. 226–245.

[9] M. Obrist, R. Bernhaupt, E. Beck, and M. Tscheligi, "Focusing on elderly: An itv usability evaluation study with eye-tracking," in *Proceedings of the 5th European Conference on Interactive TV: A Shared Experience*, ser. EuroITV'07, 2007, pp. 66–75.

[10] M. Pirker, R. Bernhaupt, and T. Mirlacher, "Investigating usability and user experience as possible entry barriers for touch interaction in the living room," in *Proceedings of the 8th International Interactive Conference on Interactive TV & Video*, ser. EuroITV '10, 2010, pp. 145–154.

[11] R. Bernhaupt, M. Boutonnet, B. Gatellier, Y. Gimenez, C. Pouchepanadin, and L. Souiba, "A set of recommendations for the control of iptv-systems via smart phones based on the understanding of users practices and needs," in *Proceedings of the 10th European Conference on Interactive Tv and Video*, ser. EuroiTV '12. New York, NY, USA: ACM, 2012, pp. 143–152.

[12] R. Bernhaupt and M. M. Pirker, "User interface guidelines for the control of interactive television systems via smart phone applications," *Behaviour & Information Technology*, no. ahead-of-print, pp. 1–16, 2013.

[13] Apple, "Apple airplay," http://www.apple.com/airplay/, 2014.

[14] P. Centieiro, T. Romão, and A. E. Dias, "Enhancing remote live sports experiences through an eyes-free interaction," in *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, ser. MobileHCI '13. New York, NY, USA: ACM, 2013, pp. 65–68.

[15] S. K. Kane, J. P. Bigham, and J. O. Wobbrock, "Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques," in *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility*, ser. Assets '08. New York, NY, USA: ACM, 2008, pp. 73–80.

[16] R. Bernhaupt, M. Boutonnet, B. Gatellier, Y. Gimenez, C. Pouchepanadin, and L. Souiba, "A set of recommendations for the control of iptv-systems via smart phones based on the understanding of users practices and needs," in *Proceedings of the 10th European Conference on Interactive Tv and Video*, ser. EuroiTV '12, 2012, pp. 143–152.

[17] J. Nielsen and R. Budiu, *Mobile Usability*. New Rider, Oct. 2012.

[18] J. Sauro, "Measuring usability with the system usability scale," http://www.measuringusability.com/sus.php, 2011.

[19] A. Bangor, P. T. Kortum, and J. T. Miller, "An empirical evaluation of the system usability scale," *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574–594, 2008.