Winter 2010

Using voice to tag digital photographs on the spot

Michael A. Farrar
University of New Hampshire, Durham

USING VOICE TO TAG DIGITAL PHOTOGRAPHS ON THE SPOT

BY

MICHAEL A. FARRAR

BS CE, University of New Hampshire, 2007

THESIS

Submitted to the University of New Hampshire

In Partial Fulfillment of

the Requirements for the Degree of

Master of Science

In

Electrical Engineering

December, 2010

This thesis has been examined and approved.

Thesis Director, Andrew L. Kun
Associate Professor of Electrical and
Computer Engineering

W. Thomas Miller, III
Professor of Electrical and Computer
Engineering

Nicholas Kirsch
Assistant Professor of Electrical and
Computer Engineering

12/14/2010

Date

ACKNOWLEDGEMENTS

I would like to thank Dr. Andrew L. Kun for the guidance and support which he offered throughout my academic curriculum, firstly as his student, but foremost, as my advisor.

I thank Dr. W. Thomas Miller, III for his technical efforts in the implementation of this work and for serving on my thesis committee.

I also thank Nicholas Kirsch for serving on my thesis committee.

Thank you to everyone involved in the Consolidated Advanced Technologies Laboratory for providing the opportunity to perform research with Project54 and pursue a graduate degree at the University of New Hampshire. It has been a pleasure working with you.

Finally, I thank my family and friends for their love and support. I would not possess the qualities of life which I greatly value without your encouragements or perspectives.

This work was supported by the U.S. DOJ, Department of Justice under grants 2005CKWX0201 and 2006DDBXK099.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF ACRYONYMS

ANOVA	...	Analysis of Variance
API	...	Application Programming Interface
CA	...	Control Action
CE	...	Embedded Compact
Fb	...	Filtered by
GUI	...	Graphical User Interface
HCI	...	Human Computer Interaction
HTC	...	High Tech Computer Corporation
HTTP	...	Hypertext Transfer Protocol
MD5	...	Message-Digest algorithm 5
P54	...	Project54
PC	...	Personal Computer
PDA	...	Personal Digital Assistant
QWERTY	...	Definition: Modern-day English-language keyboard layout
REST	...	Representation State Transfer
SDK	...	Software Development Kit
SDP	...	Single Day Participants
SUI	...	Speech User Interface
T9	...	Text on 9 keys

TE	...	Tagging Event
TEP	...	Tag Existence Percentage
TIP	...	Tag Insertion Percentage
URL	...	Uniform Resource Locator
wpm	...	Word per minute
XML	...	Extensible Markup Language
XP	...	eXPerience

ABSTRACT

USING VOICE TO TAG DIGITAL PHOTOGRAPHS ON THE SPOT

By

Michael A. Farrar

University of New Hampshire, December, 2010

Tagging of media, particularly digital photographs, has become a very popular and efficient means of organizing material on the internet and on personal computers. Tagging, though, is normally accomplished long after the images have been captured, possibly at the expense of in-the-moment information. Although some digital cameras have begun to automatically populate the various fields of a photograph's metadata, these generic labels often lack in the descriptiveness presented through user-observed annotations and therefore stress the necessity of a user-driven input method. However, most mobile annotation applications demand a great number of keystrokes in order for users to tag photographs and thereby focus the user's attention inward. Specifically, the problem is that these applications require users to take their eyes off the environment while typing in tags. We hypothesize that we can shift the user's focus away from the mobile device and back to their environment by creating a mobile annotation application which accepts voice commands. In other words, our major hypothesis is that a convenient way of tagging digital photographs is by using voice commands.

CHAPTER 1

INTRODUCTION

1.1 Tagging of digital photographs: problem statement

Tagging of media, particularly digital photographs, has become a very popular and efficient means of organizing material on the internet and on personal computers. Over a short period of time the technique has evolved from an optional feature to a must-provide service and can be found within modern desktop and internet photo galleries. However, tagging is normally accomplished long after the images have been captured, possibly at the expense of in-the-moment information. Although some digital cameras have begun to automatically populate the various fields of a photograph's metadata, typically including date and time of capture and details of camera settings such as focal length, aperture and exposure, these generic labels often lack in the descriptiveness presented through user-observed annotations.

A typical tagging scenario is depicted in figure 1.1 which demonstrates the use of Windows Live Photo Gallery. Prior to tagging, users must invoke the application's import wizard, shown in the leftmost portion of the figure, where they are allowed to make selections specifying the photographs to be transferred from their digital camera device. With transfer complete, tags may be applied to individual photographs or in a group-like manner as shown in the rightmost portion of the figure. This digital camera-personal

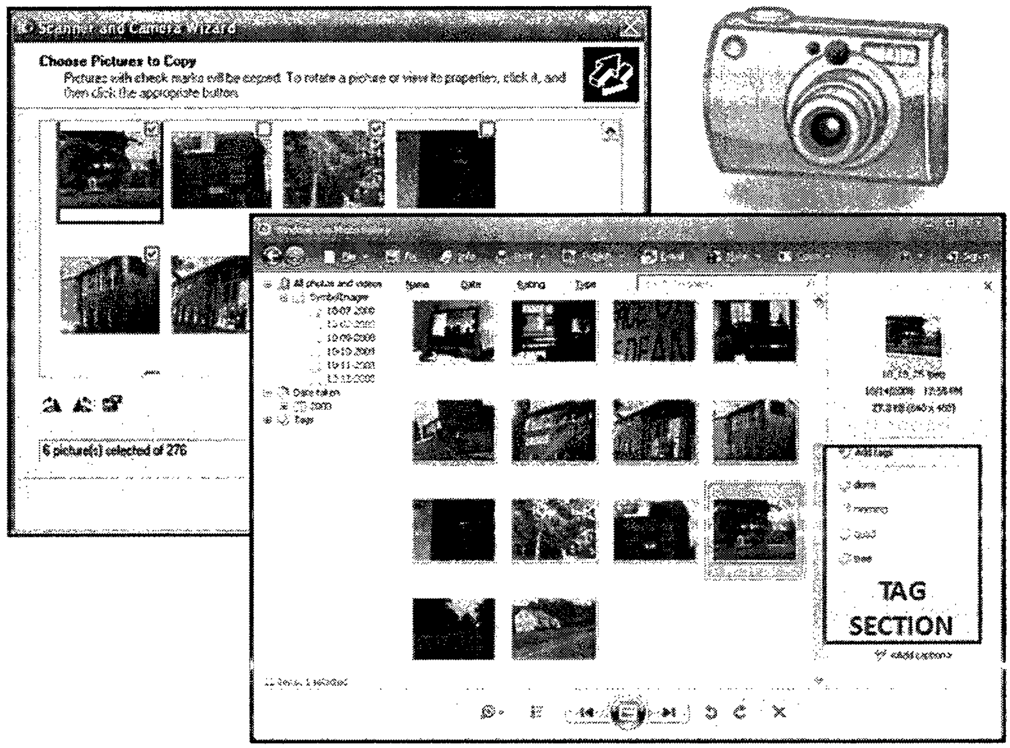


Figure 1.1. Typical tagging scenario of Windows Live Photo Gallery

computer relationship limits a tagging application's range of mobility, subsequently imposing a greater time lapse from moment of capture to moment of annotation.

Today's typical digital camera is the child of a technology-replacing-technology movement, a shift from uni-functional devices to multipurpose cellular phone-like devices such as the Pocket PC. Advances in optics, communication and fabrication supply consumers with pleasing photography and seamless internet access in a small form-factor package, allowing applications such as ZoneTag [1] and Shozu [2] to initiate the "unplugged" world of annotation. Further, these applications have begun to incorporate the services of online photo-managing and -sharing websites like Flickr [3], revolutionizing photography with their immediate upload implementations. However, the annotation processes of most mobile annotation applications often demand a great number of keystrokes in order to tag a photograph, thereby focusing the user's attention inward. Specifically, the **problem** is that these applications require users to take their eyes off the environment while typing in tags.

1.2 Goals

Our **primary goal** is to explore methods of tagging photographs which allow users to focus on their environment. We break this general goal into two sub goals. Our **first sub-goal** is to create, test and deploy a mobile application which will allow users to tag photographs without drawing their immediate attention for extended periods of time. Once accomplished, this will allow us to pursue our **second sub-goal**, which is to create, test and deploy a flexible software infrastructure allowing the collection of data regarding human-computer interactions (HCI), photographs captured, etc. generated through the use of our tagging application.

1.3 Hypotheses

We hypothesize that we can accomplish our first goal by enabling the application to accept voice commands. In other words, our **major hypothesis** is that a convenient way of tagging digital photographs is by using voice commands, thereby focusing the user's attention outward on the environment and possibly resulting in enriched observations (in comparison to the observations made when tagging photographs at a later time).

When using command-and-control speech interaction (voice commands) one must provide the speech recognizer with a grammar. Entries of this grammar define the validity of user utterances; meaning, the success of a voice command depends upon the existence of the intended command in the grammar. Our **second hypothesis** is related to the initialization of this grammar, where we hypothesize that in order to allow users to tag photographs using speech we should populate the grammar with user-dependent information obtained from two online resources:

1. *Tags generated by the user and attached to photographs on one or more of the user's online photo-sharing accounts.* As users tag photographs, they are likely to want to tag many of them with the names of family members, friends and pets, visited locations, etc. (figure 1.3a left). While one way to retrieve this information for initializing the grammar would be to simply ask the users, we expect that for those who have already started tagging photographs online, a quicker and more convenient way would be to query their accounts.
2. *Tags related to particular interests of the user generated by other users with similar interests.* We also expect that users will often take sequences of

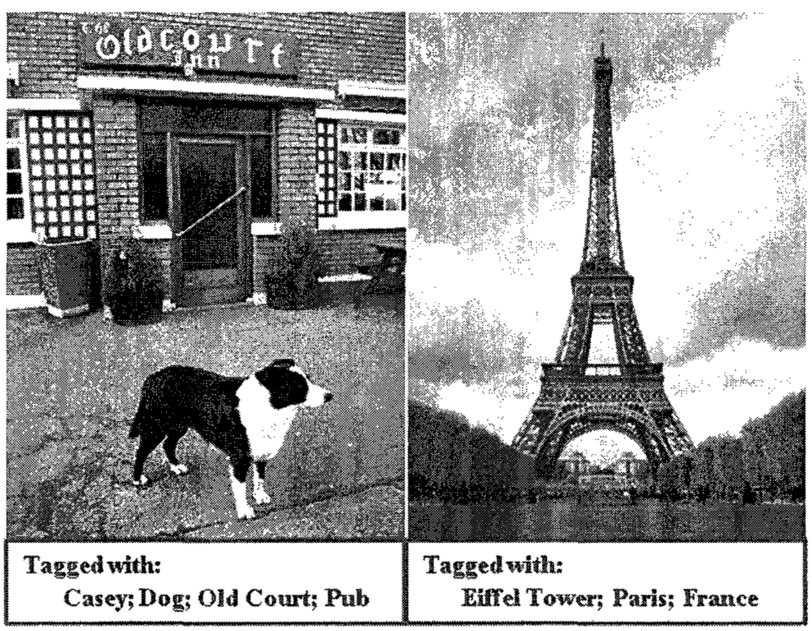


Figure 1.3a. Photographs demonstrating our user-dependent approach to grammar initialization

photographs that are common in subject (e.g. a trip to the ball park or to Paris) and that many of the tags paired with these images will be similar to those used by others with similar interests (figure 1.3a right).

Of course, many people are quite comfortable using the keyboards of mobile phones. Our **third hypothesis** is that self-identified highly-experienced keyboard users, because of their increased experience with the keyboard, will use voice commands less than those with lower keyboard experience, especially if they have only a short period of time (e.g. one day) to adjust to the voice command input method of our tagging application. Figure 1.3b demonstrates the differences in the text entry processes of our mobile devices, showing the T9-equipped HTC Touch Diamond in the leftmost portion of the figure and the QWERTY-equipped HTC Touch Pro in the rightmost portion of the figure. As an additional measure of control, and in order to evaluate differences in the

Figure 1.3b. A depiction of the differences in the text entry processes of our T9-equipped (left) and QWERTY-equipped (right) devices

number of tags generated between keyboard types, we intend to counterbalance keyboard type among our male and female participants, **hypothesizing** that those suited with the T9-equipped device will generate a greater number of tags over those suited with the QWERTY-equipped device since the text entry process on our T9 device avoids sliding out the keyboard.

For our fifth **hypothesis**, we hypothesize a correlation between voice command usage and task completion rate; that is, users who unsuccessfully issue voice commands relatively early in the process of learning how to use our tagging application will not keep at it, they will give up on voice commands and pursue other interaction modalities. One reason for a poor task completion rate on any speech-enabled device is the occurrence of out-of-grammar utterances. These voice commands are unrecognizable by the application since they are not explicitly included in the grammar.

1.4 Approach

We divide this research into four phases: (1) development of the mobile tagging application, (2) development of the data collection server, (3) experiment with human participants and (4) data analysis.

1.4.1 Development of the mobile tagging application

A constraint of this phase is our desire to use Windows Mobile 6.0/Windows CE 5.0 devices, allowing us to take advantage of our laboratory's experience with the Windows operating system. The graphical (GUI) and speech user interface (SUI) of the tagging application are to be structured through Project54 (P54) libraries [4] and implemented under the Microsoft Visual Studio 2005 development environment. Various

Windows software development kit (SDK) procedure calls will provide the tagging application with camera functionality. Similar to some existing mobile-annotation applications, and to evaluate our hypothesis on grammar initialization, this work will incorporate the services of Flickr. Communication with Flickr is achieved utilizing their application programming interface (API) [5]. Figure 1.4.1 depicts an abstract of our application's design, keeping in mind that we must preserve the simplistic qualities of the digital camera and the tagging process in implementing each of our application's software components. The "Talk" button, highlighted in red, will be used to activate P54's speech input library, which utilizes the framework of Microsoft Voice Command in the decoding of user utterances (handled on-device).

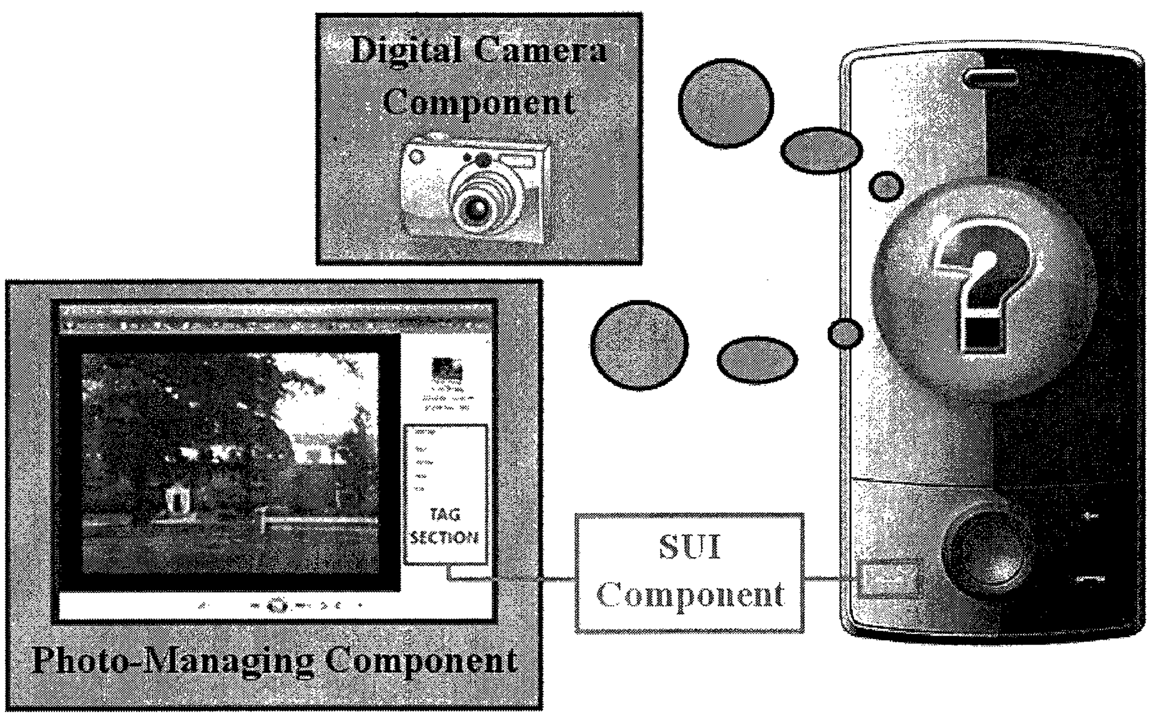


Figure 1.4.1. Abstract of our application's design targeting the Windows Mobile 6.0 platform

1.4.2 Development of the data collection server

A constraint of this phase is our desire to use Windows XP desktop machines, again allowing us to take advantage of our laboratory's experience with the Windows operating system. The P54 server architecture [6] will be extended to support the demands of this research. All pertinent data is to be logged via cellular connection.

1.4.3 Experiment with human participants

We propose a between-subjects experiment with counterbalanced QWERTY/T9 device usage among male and female participants. Participants will extend the activities found in the use of a typical digital camera by tagging the photographs they capture. There will be no emphasis on which method of tagging a participant should pursue (GUI or SUI).

1.4.4 Data analysis

Both qualitative and quantitative methods will be used to analyze participants. The qualitative analysis will include pre- and post-experiment Likert scale-based [7] questionnaires concerning the participant's level of familiarity with digital camera usage, their understanding of the concept of tagging photographs and general questioning on the performance of the application and modality preference. The quantitative analysis will categorize each participant's usage of the application based on the information found within their logging files. These files will contain all interactions with the application's interfaces.

CHAPTER 2

BACKGROUND

2.1 The mobile computing timeline

Growing in their acceptance, mobile devices have begun to weave their way into the everyday lives of more and more individuals. The era of the standalone personal computer running simple process input-produce output applications is evolving into a sophisticated ubiquitous computing network where calculation of "the answer" requires multiple information sources. Accompanying this complex architecture are the countless reductions in physical dimensions seen throughout such devices' timelines. Unquestionably, these modifications were necessary in order to extend mobility into the various domains of today's technological world; however, each mark only broadens the gap between device size and HCI methods. The much-sought-after small form-factor, when improperly equipped, places restrictions upon users which often result in frustration growth and device performance degradation. Following the trend of computers in general, where an existing configuration becomes outdated quite rapidly, then soon again these timelines will initiate our corrective efforts, thus revealing the endless complications faced in mobile engineering.

2.2 The ubiquitous computing vision

What makes a technology mobile? The answer to this question is not as simple as it may seem. Of course size, source and consumption of energy, computational speed and flexibility are all crucial attributes of mobility; yet, their dependencies usually draw some amount of user compensation (e.g. the dependency of computational speed upon physical size requires compensation through runtime). Weiser [8;9] and Brown [9] envision a world where systems, not users, adapt to meet demands through a mixture of embodied virtuality and calm technology, bringing computer-readable data into the physical world without overburdening users. They argue the necessity of a user-selective presentation of information dispersed over central and peripheral attention foci, claiming to impose less frustration and greater environmental awareness through an offer-but-not-demand level of disruption.

2.3 Connectedness: user-to-user and user-to-information

Mobile devices are highly personal, and regarding the presentation of data, highly customized to their user's preferences, often leaning towards more computationally and energy expensive formats. Before addressing the HCI issues of the introductory paragraph (section 2.1), we must first understand these modernized presentations and their areas of applicability. Text and graphics such as pictures, photographs and animations have cooperatively captured the screen and momentarily fulfilled market demands. People want to be aware of one another and of information. Whether it be through email or the instant message, user-to-user connectedness via textual/graphical exchange is becoming, if it is not already, the most popular technique to satisfying awareness.

Periodically, as technologies evolve, a culture will witness the merging of two data representations, in this case text and graphics. The works of Kindberg et al. [10], Van House [11] and Farkas et al. [12] set to clarify the true value of modern photography by examining its life-cycle of activity. Summarizing their results in the words of Kindberg et al., use of the camera phone may be classified under an {[affective, functional] x [social, individual]} taxonomy matrix of reasons for image capture. Kindberg et al. conclude that the camera phone, itself, is most often used for personal reflection experiences; however, and in slight contradiction, Ames and Naaman [13] and Marlow et al. [14] find the annotation of images to be a social-organizational form of user-to-user connectedness.

> *"The camera phone is neither an incremental step forward from a mobile phone, nor a poor relation of a digital camera. Rather, it is a device which is sometimes used like a digital camera, but is different in the range of activities it supports."*
>
> – Kindberg et al. [10]

Marlow et al. [14], Tomas et al. [15] and Topkara et al. [16] explore the vocabulary problem, stated by Marlow et al. as the occurrence of different users using different terms to describe the same things. In their work, Marlow et al. evaluate the annotation characteristics of a subset of photos from the Flickr community, finding the strongest relationship between variables in the set [photos, distinct tags, contacts] to be between photos and distinct tags (an increase/decrease in photos implied an increase/decrease in a user's distinct tags). This result, and the lacking relationship between contacts and distinct tags, suggests the tagging of one's own photographs to be the dominant form of tag generation. It was noted, however, that most users had very few distinct tags and that much of their corpus was in overlap with the community.

Marshall and Brush [17] define three annotation types which characterize the relationship between the personal annotations people make while reading and the annotations they share when discussing the same materials online. They conclude that the majority of personal annotations are anchor-only markings, an annotation similar to tags, which undergo a "nothing-to-something" or "original-and-more" content change prior to being shared; in other words, the annotating individual had been reminded of, and was able to expand upon, their in-the-moment topic-of-interest from these "tags" and that this information could then be used to relay subjective matter socially. Unquestionably, this result strengthens the preservation of in-the-moment information via tagging captured photographs, hopefully positioning the viewer in the author's state-of-mind.

2.4 A focus on user interaction methods

2.4.1 Tactile input

An expanding line of input mechanisms are available internally and externally to the mobile device. Full-sized mechanical keyboards have claimed their fair share of surface space along with their graphical keyboard counterpart which occludes an extreme percentage of the display while active. Another, and perhaps more considerable, issue with such input styles can be found in their interaction times, which suffer greatly when compared to desktop mechanical keyboards. Even further, most mobile environments are unsuitable for mouse operation, and the accurate selection-by-tracking it provides has been replaced by direct-touch methods invoked by finger, thumb or stylus. Although the stylus is significantly more accurate than finger or thumb, users often forget about it or feel that it takes too much time and effort to retrieve.

Consumers' growing desire for information has in essence redirected the content of the desktop display to the mobile display, ensuing in rich interfaces composed of targets too small to be selected reliably through touch. As just mentioned, many users are reluctant to use a stylus, leaving tactile selections as their preferred method of interaction. Therefore, an assessment must be made of when direct-touch becomes an acceptable form of input. The works of Vogel and Baudisch [18] and Karlson and Bederson [19] make this assessment with target size conditions of 10.4 mm^2 or greater yielding selection error rates of at most 5%, as confirmed by Bogel and Baudisch. Contrarily, as target sizes are reduced Bogel and Baudisch find error rates to increase exponentially. Although these findings were also dependent upon the number of targets, their locations and orientations, from an interactive standpoint it would seem as though the burden of direct-touch outweighs its benefit.

As debated by Yatani and Truong [20], most text entry techniques only support input via the user's dominant hand; and, most of these techniques have been evaluated only while the user is stationary. In their analysis, Yatani and Truong reveal the impacts of mobility upon text entry with respect to speed, accuracy and mental workload, finding that input speed generally decreases while error rate and mental workload increase when users are presented with an in-motion task.

2.4.2 Speech input: towards a multimodal interface

James and Reischel [21] compare the word-per-minute (wpm) input rates for multitap, T9, desktop-mechanical and SUI mechanisms, stating an expert-user rate of 11 wpm, 26 wpm, 80 wpm and 200 wpm, respectively. The environments of mobile devices, though, are very dynamic, and often pose problems across a number of input methods. Reductions in the signal-to-noise ratio of a speech input will dramatically influence the accuracy of even the best speech recognition engines, potentially reducing their 200 wpm input rate to something much less attractive. Turner and Kun's [22] analysis of the P54 command-and-control SUI was found to peak with accuracy of 94.02% and average at 85.34%; overcoming the noisy in-car environment of the police cruiser using noise-canceling microphones and static grammar files which held the computer's number of utterance comparisons to a fixed value. Their use of the GUI as an "always open" window to the grammar's contents, where graphical button labels corresponded to SUI commands, provided officer's a strong assurance to the state of the recognition engine and stabilized SUI performance.

Considering the advantages of the SUI, Cox et al. [23] explore the viability of speech recognition as an alternative method to text entry, not with the intention of replacing the traditional keypress mode of interaction, but instead to add functionality to the interface. In their work they consider two navigation modalities in conjunction with three message editing modes, forming the interface manipulation matrix {[keypress, speech] x [multitap, T9, speech]}. Results of the study show that conditions using spoken text entry produce the fastest task completion times and that participants preferred spoken interaction over tactile, perceiving it to require less effort. It is worth noting, however, that user preferences did not entirely agree with performance. A unimodal speech interaction was much preferred whereas a mixed mode of keypress navigation and speech text entry actually gave better performance. In general, this result seems to suggest that users are prepared to sacrifice task completion and accuracy in favor of a less demanding interface and do not feel any strong need for keypress text entry.

"The choice of input modes may be influenced by individual preference or the nature of the surrounding environment. For example, sending a voice text may not be the most appropriate method in a public place whereas spoken input may be the preferred option in a mobile environment or when simultaneously performing another task. Thus, offering a choice of input modalities suggests that a wider range of users, tasks and environments can potentially be accommodated." – Cox et al. [23]

2.5 Efficiency

Efficiency, the term itself can be molded to represent an almost infinite number of domains; and thus far, every topic and subtopic in one way or another, whether it be used in relation to the productivity of the user or the system, could be seen as a fundamental branch of the expression. For instance, the explorations detailed in sections 2.3 and 2.4 contribute to efficiency by increasing a user's task completion rate, making the user more efficient at distributing, gathering and inputting data. Efficiency, then, is analogous to mobility.

2.5.1 Context as an efficiency

Context, as defined by Dey and Abowd [24], is any information that can be used to characterize the situation of a person, place or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves. A context-aware system, they claim, is one which uses context to provide the user with relevant information and services. Knowing a user's location within a network, then, becomes an extremely important attribute to the success of mobility in general, stated by Dey and Abowd as being one of the primary entities of context. On the other hand, under what circumstances is physical location still a meaningful ingredient of context? What levels of precision should be sought? In their evaluation of location-aware computing, Hazas et al. [25] divide location-sensing technologies into coarse- and fine-grained systems; and Pfeiffer [26] underlines that in order for location-aware computing to become part of everyday life, devices must be enabled to flex between the two technology groupings (between course- and fine-grained location sensing technologies), benefiting from one when the other is unavailable – a schematic still in its developmental stages.

Ames and Naaman [13] and Sigurbjornsson and Zwol [27] explore two very different annotation recommendation strategies, location-based and collective knowledge-based, respectively. The latter analysis classifies a subset of the Flickr community, revealing a broad semantic spectrum focused in the set [where, who, what, when], while Ames and Naaman compare and contrast two camera phone-based tagging applications, ZoneTag [1] and Shozu [2], and conclude with these guidelines-to-design:

1. Make the annotation pervasive and multi-functional
2. Make it easy to annotate when the information is captured
3. Do not force users to annotate at the point of information capture
4. For systems that have both mobile capture and desktop- or web-based components, allow annotation in both settings
5. Relevant tag suggestions, even when not used directly, can encourage tagging and give users ideas about possible tags

2.5.2 An alternative view

Rogers [28] reconsiders the passive environment put forth by peer researchers, claiming a shift from proactive computing to proactive people as being more beneficial to society.

> *"The specifics of the context surrounding people's lives are much more subtle, fluid and idiosyncratic than believed, making it difficult, if not impossible, to implement context in any practical sense."* – Rogers [28]

Computers were designed as tools, as devices and as systems which extend and engage people in their pursuits. Why then should this excitement of interaction be avoided? Erickson [29] hypothesizes the realization of a truly context-aware system by allowing humans to evaluate the data of computational systems. As is today, a system's comprehension of context is represented through a relatively small set of quantitative variations; and in comparison, a human's comprehension of context is built upon observations derived from both obvious and subtle cues weighted through past experiences. Understanding this neurology, Erickson elaborates on the misuse of the term "context-aware" claiming it to invoke powerful notions of "context" and "awareness", concepts which people understand very differently from the way in which they are being instantiated in context-aware systems.

2.6 Our focus

The drives of photography are changing from personal reflection (Kindberg et al. [10]) to social-organizational (Ames and Naaman [13] and Marlow et al. [14]). Revelations similar to Van House's [11] of Flickr being seen as a social site where image collections are perceived as streams, not archives, now emphasize the tagging process as a whole. Photographs have become multipurpose; therefore, the applications which manipulate them must become multipurpose. Yatani and Truong's [20] study on the impact of mobility upon text entry with respect to speed, accuracy and mental workload, and the performance measurements of Turner and Kun [22] and Cox et al. [23], urge a movement towards multimodal interactions.

In order to preserve the simplistic qualities of the digital camera and the tagging process it then becomes important to understand the guidelines-to-design presented by Ames and Naaman [13] and their underlying relationship to the conceptual ideas put forth by Weiser [8;9] and Brown [9]. The methods of our tagging application greatly depend upon the notion of a tag bank: a stored listing of textual labels composing the grammar and largely defining the validity of user utterances. We make the following argument in regards to initialization of the grammar:

1. Dey and Abowd [24] define a context-aware system as one which uses context to provide the user with relevant information and services.
2. Sigurbjornsson and Zwol's [27] classification of a subset of the Flickr community reveal a broad semantic spectrum focused in the set [where, who, what, when], where the entities of [who] and [what] may require, as portrayed by Pfeiffer [26], Rogers [28] and Erickson [29], unavailable contextual knowledge.
3. The annotation characteristics evaluated by Marlow et al. [14], again over a subset of the Flickr community, suggest the tagging of one's own photographs to be the dominant form of tag generation.

The tag bank, then, is to be populated proactively.

CHAPTER 3

APPROACH

3.1 Phases of the research

As previously stated, this work will explore the benefits of tagging digital photographs upon capture. Our primary goal is to explore methods of tagging which allow users to focus on their environment. We hypothesize that a convenient way of tagging digital photographs is by using voice commands, thereby focusing the user's attention outward on the environment and possibly resulting in enriched observations (in comparison to the observations made when tagging photographs at a later time). We divide this research into four phases:

1. Development of the mobile tagging application
2. Development of the data collection server
3. Experiment with human participants
4. Data analysis

3.2 Development of the mobile tagging application

3.2.1 Application overview

The GUI of our tagging application consists of six graphical windows and an external camera utility: the First Use window, the Manager window, the external camera utility, the Tags window, the Tag Bank window, the More window and the Upload window. The hierarchical diagram of figure 3.2.1 depicts these windows' orientations

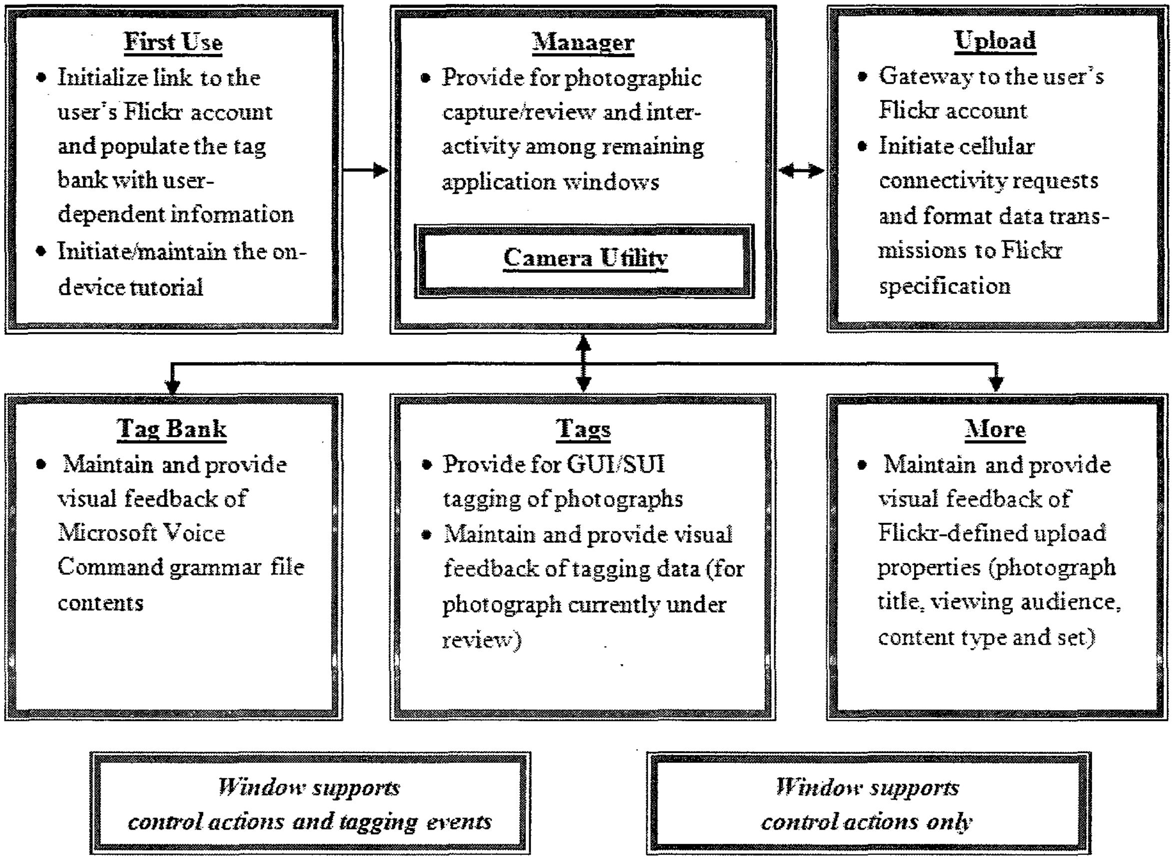


Figure 3.2.1. Window hierarchy showing control action and tagging event support

within the layout of our application. To be discussed later in the work, we term all GUI/SUI interactions in which the user traverses the application's tree-like structure or accesses some feature of the application not related to tagging data as control actions. All GUI/SUI interactions which manipulate tagging data are termed tagging events.

The development platform was chosen to be generic across all Windows Mobile 6.0/Windows CE 5.0 devices, each implementing a manufacturer-specific routine providing camera functionality through SDK procedure calls. Standard Microsoft Windows creation and notification handlers are structured through P54 GUI libraries implemented under the Microsoft Visual Studio 2005 development environment. Along with this GUI structure, P54 utilizes the framework of Microsoft Voice Command in the decoding of user utterances.

Similar to some existing mobile-annotation applications, this work incorporates the services of the online photo-managing and -sharing website Flickr. Communication with Flickr is achieved via REST-HTTP GET/POST actions to the endpoint URL: http://api.flickr.com/services/rest/. Table 3.2.1 depicts a sample REST-HTTP GET/POST action outlining the [flickr.auth.getFullToken] method from subsection 3.2.2. As outlined in the table, each method requires a select number of parameters. If the method is signed (secure), a signature parameter (the [api_sig] parameter) is also required, which is computed as the MD5 sum of the alphabetical ordering of the method's name and its arguments, all in string format. Using the REST-HTTP POST action with calling convention similar to that from the table produces an XML response containing the desired information; in this case, the user's full account token and identification number. For additional information on the use of Flickr's API see Flickr's services URL: http://www.flickr.com/services/api.

Method Name	**Parameters**	**XML Response Parsers**
flickr.auth.getFullToken	api_key mini_token api_sig	"<token>" "<user_nsid>"

Method Calling Convention
http://api.flickr.com/services/rest/ ?method=flickr.auth.getFullToken &api_key=<api_key> &mini_token=<mini_token> &api_sig=<api_sig>

Table 3.2.1. REST-HTTP GET/POST actions outline for the [flickr.auth.getFullToken] method

3.2.2 The First Use window

As users tag photographs, they are likely to want to tag many of them with the names of family members, friends and pets, visited locations, etc. While one way to retrieve this information for initializing the grammar would be to simply ask the users, we expect that for those who have already started tagging photographs online, a quicker and more convenient way is the querying of their account. Figure 3.2.2 depicts the First Use window in its state following the user code retrieval processes of the "Login", "Consent" and "User Code" Internet Explorer windows. This sequence of activity is a requirement of the Flickr API and issues our application rights to access information from the user's account. Once the link to their account has been established, the First Use window automates a tag bank initialization routine where 100 of the user's most frequently used tags are extracted from his or her account and inserted into the tag bank for later use.

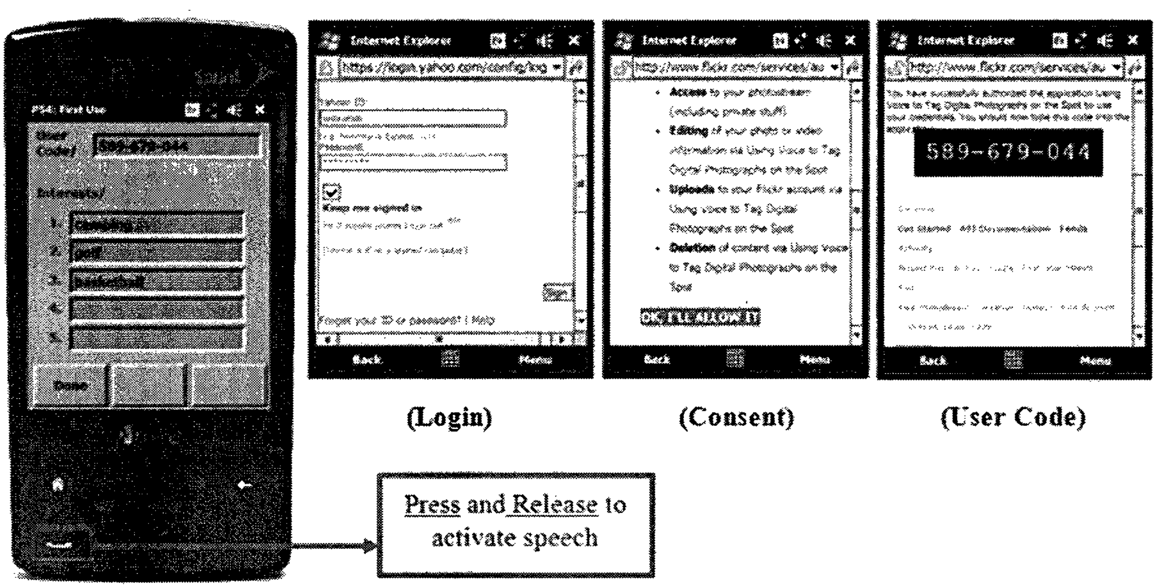


Figure 3.2.2. The First Use window: linking with Flickr and initializing the tag bank

We also expect users to take sequences of photographs that are common in subject (e.g. a trip to the ball park or to Paris) and that many of the tags paired with these images will be similar to those used by others with similar interests. To maximize the versatility of the tag bank, the First Use window allows for the user to enter up to five photographic interests as demonstrated in figure 3.2.2 with the interests of "camping", "golf" and "basketball". Each interest is compared against Flickr's tag database for similarities, the results of which are inserted into the tag bank for later use at a limiting rate of 25 tags per interest. In combination, the two initialization routines generate a maximum of 225 tag bank entries and form a basis for the evaluation of our second hypothesis on grammar initialization. Table 3.2.2a summarizes the API methods used in performing the above initialization actions.

Method Name	Parameters	XML Response Parsers
flickr.auth.getFullToken	api_key mini_token api_sig	"<token>" "<user_nsid>"
flickr.tags.getListUserPopular	api_key auth_token user_id count api_sig	"<tag count>" "<tag>"
Flickr.tags.getClusters	api_key tag	"<tag>"

Table 3.2.2a. REST-HTTP GET/POST actions outline for the First Use window

As with all windows of our application, the First Use window accepts voice commands as input. These voice commands and their graphical counterparts are defined in table 3.2.2b. To activate the recognition engine the "Talk" button of the device's keypad must be pressed as shown in figure 3.2.2.

Graphical Button	Voice Command	Function
Done	"Done"	Exits the First Use window

Table 3.2.2b. Graphical and voice commands of the First Use window

3.2.3 The Manager window and external camera utility

The Manager window, depicted in the leftmost portion of figure 3.2.3, is the main window of our application and allows for the user to view the on-device tutorial, upload previously captured photographs to Flickr, manipulate previously captured photographs, access the tagging windows and capture new photographs. Also shown in the figure is the HTC Touch version of the Windows Mobile 6.0 camera utility. Working in cooperation, the two applications are capable of exchanging high-resolution photographs while preserving their original capture orientations of landscape or portrait, precisely simulating the digital camera interface. Table 3.2.3 defines the Manager window's voice commands and their graphical counterparts; however, it should be noted that voice commands are not

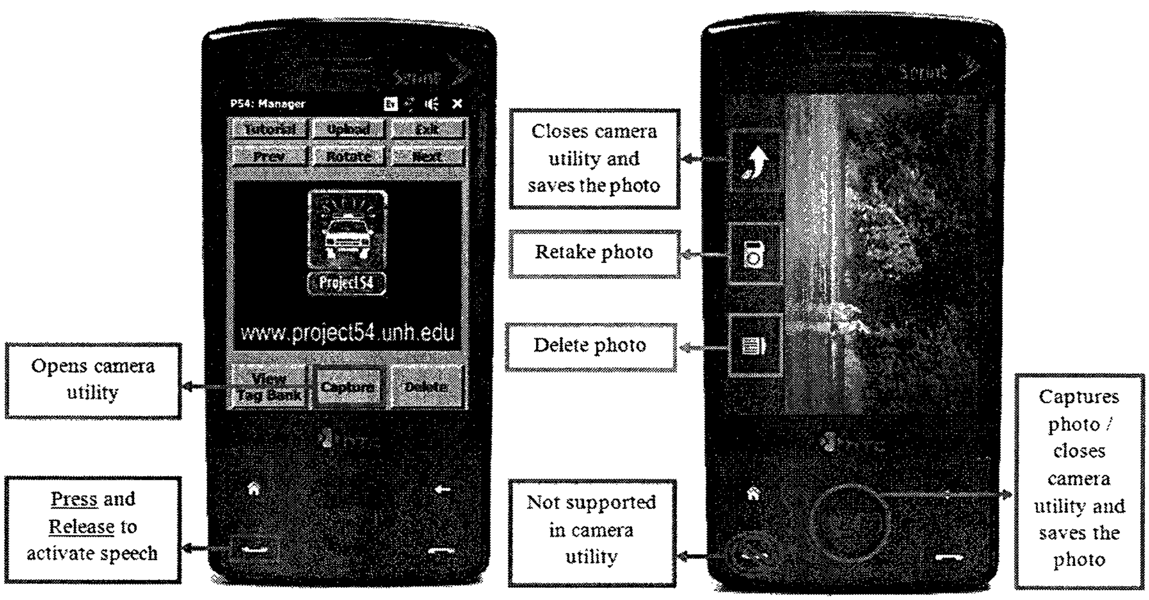


Figure 3.2.3. The Manager window (left) and external camera utility (right)

supported by the camera utility and that the "Action" button of the device's keypad must be used to capture photographs.

Graphical Button	Voice Command	Function
Tutorial		Opens the on-device tutorial
Upload	"Upload"	Opens the Upload window
Exit		Exits the tagging application
Prev	"Prev" or "Previous"	Scrolls photographs in an earlier-date manner
Rotate	"Rotate"	Rotates a photograph counterclockwise
Next	"Next"	Scrolls photographs in a latter-date manner
View Tag Bank / Tags	"View Tag Bank" / "Tags"	Opens the Tag Bank / Tags window
Capture / Done	"Capture" / "Done"	Opens / Closes the external camera utility
Delete		Deletes the current photograph

Table 3.2.3. Graphical and voice commands of the Manager window

3.2.4 The Tags window

Immediately upon capture, or by reviewing previously captured photographs, the [Tags] command is made available by the Manager window. Selecting it will open the Tags window, depicted in figure 3.2.4, where the photograph may be tagged in one of three ways: via (1) text entry, (2) graphical selections or (3) voice commands. These operations are summarized

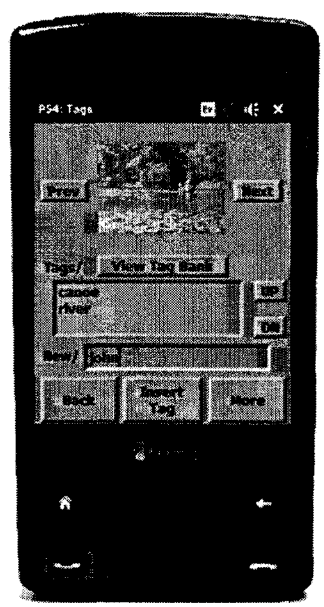


Figure 3.2.4. The Tags window

below. Table 3.2.4a defines the Tags window's voice commands and their graphical counterparts.

1. Text entry – New tags may be typed into the new text field, as demonstrated in the figure with the tag "john", and inserted using the [Insert Tag] command. Newly typed tags are paired with the photograph and stored in the tag bank.
2. Graphical selections – Tags may be removed from the photograph by selecting them with the stylus. If an incorrect selection has been made the tag may be reinserted manually or by using the "Undo" voice command of table 3.2.4b.

3a. Removal voice commands – Tags may be removed from the photograph using the tagging voice commands of the Tags window. These commands are listed in table 3.2.4b.

3b. Copying voice commands – Some mobile annotation applications demand a great number of keystrokes in order to tag a photograph. We believe the act of typing to focus the user's attention inward, off the environment and onto the device. A more convenient method of tagging digital photographs, we hypothesize, is through the use of copying voice commands, allowing the user to refocus on their environment. Tags may be copied to a photograph from the tag bank using the tagging voice commands of the Tags window. These commands are listed in table 3.2.4b. The tag bank need not be visible in order to use voice commands to pair a tag with a photograph; however, the tag must appear in the tag bank for the command to properly be recognized.

Graphical Button	Voice Command	Function
Prev	"Prev" or "Previous"	Scrolls photographs in an earlier-date manner
Next	"Next"	Scrolls photographs in an latter-date manner
View Tag Bank	"View Tag Bank"	Opens the Tag Bank window
UP	"Up" or "Scroll Up"	Scrolls tags in an upward manner
DN	"Down" or "Scroll Down"	Scrolls tags in a downward manner
Back	"Back"	Returns to the previous window
Insert Tag	"Insert Tag"	Pairs a newly typed tag with the photograph
More	"More"	Opens the More window

Table 3.2.4a. Graphical and voice commands of the Tags window

Voice Command	Function
"Remove Tag *<tag>*"	Removes the tag specified by *<tag>* from the photograph
"*<tag>*"	Copies the tag specified by *<tag>* from the tag bank to the photograph
"Undo"	Removes the lastly added or reassigns the lastly removed tag to the photograph

Table 3.2.4b. Tagging voice commands of the Tags window

Considering some examples:

Ex 3.2.4a. Saying "Remove Tag river" will remove the tag of "river" from the photograph of figure 3.2.4.

Ex 3.2.4b. Following example Ex 3.2.4a, saying "river" will copy the tag of "river" from the tag bank (figure 3.2.5) to the photograph of figure 3.2.4.

Ex 3.2.4c. Following example Ex 3.2.4b, saying “Undo” will remove the tag of “river” from the photograph of figure 3.2.4.

For simplicity, tags are stored in text files titled with the photographs’ time-of-capture (*<timestamp>*_Tags.txt). As each photograph is reviewed, the tags are pulled from its corresponding text file and displayed to the user.

3.2.5 The Tag Bank window

The Tag Bank window, shown in figure 3.2.5, is accessible from both the Manager and Tags windows by invoking the [View Tag Bank] command. Entries of the tag bank compose the grammar and largely define the validity of a user’s utterance; in other words, the success of a tagging voice command depends upon the existence of the intended tag in the grammar. In an effort to increase recognition performance, we’ve allow the tag bank to be edited as desired by the user (a smaller grammar imposes less stress upon the recognition engine by reducing the number of comparisons it must make for user utterances). The tag bank, then, has two modes of operation: (1) normal and (2) editing.

In its normal mode of operation, the photograph of figure 3.2.4 may be tagged from the Tag Bank window in one of three ways: via (1) text entry, (2) graphical selections or (3) voice commands. These operations are summarized below. Table 3.2.5a defines

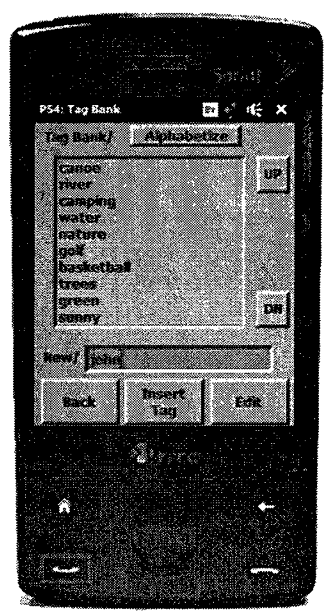


Figure 3.2.5. The Tag Bank window

the Tag Bank window's voice commands and their graphical counterparts.

1. Text entry – New tags may be typed into the new text field, as demonstrated in figure 3.2.5 with the tag "john", and inserted using the [Insert Tag] command. Newly typed tags are paired with the photograph under review (the photograph of figure 3.2.4) and stored in the tag bank.
2. Graphical selections – Tags may be copied to the photograph under review by selecting them with the stylus. To avoid a lack in visual feedback, a temporary notification is displayed upon graphically selecting a tag. If an incorrect selection has been made, the tag may be removed using the aforementioned operations of the Tags window or by using the "Undo" voice command of table 3.2.5b.
3. Copying voice commands – Tags may be copied to the photograph under review using the tagging voice commands of the Tag Bank window. These commands are listed in table 3.2.5b. The tag bank must be in its normal mode of operation in order for a copying voice command to succeed. Again, we believe the more convenient method of tagging digital photographs to be through the use of copying voice commands.

Graphical Button	Voice Command	Function
Alphabetize / Sort By Uses	"Alphabetize" / "Sort By Uses"	Reorganizes the tag bank
UP	"Up" or "Scroll Up"	Scrolls tag bank in an upward manner
DN	"Down" or "Scroll Down"	Scrolls tag bank in a downward manner

Table 3.2.5a. Graphical and voice commands of the Tag Bank window

Graphical Button	Voice Command	Function
Back	"Back"	Returns to the previous window
Insert Tag	"Insert Tag"	Pairs a newly typed tag with the photograph
Edit / Done	"Edit" / "Done"	Toggles between editing / normal modes

Table 3.2.5a (continued). Graphical and voice commands of the Tag Bank window

Some commands of the Tag Bank window are dependent upon its mode of operation. For instance, the [Insert Tag] command from table 3.2.5a, while in editing mode, will not pair newly typed tags with the photograph under review; however, these tags will be stored in the tag bank. The tagging voice commands of table 3.2.5b also share this dependency as detailed below.

Voice Command	Mode	Function
"Remove Tag *<tag>*"	Editing	Removes the tag specified by *<tag>* from the tag bank
	Normal	
"*<tag>*"	Editing	
	Normal	Copies the tag specified by *<tag>* from the tag bank to the photograph
"Undo"	Editing	Removes the lastly added or reinserts the lastly removed tag to the tag bank
	Normal	Removes the lastly added or reassigns the lastly removed tag to the photograph

Table 3.2.5b. Tagging voice commands of the Tag Bank window

Similar to the storing of a photograph's tags, the tags of the tag bank are stored in a text file (TagBank.txt). Within this file, each tag is paired with a usage marker so that the tag bank may be sorted by the tags' frequency of use if the user so chooses.

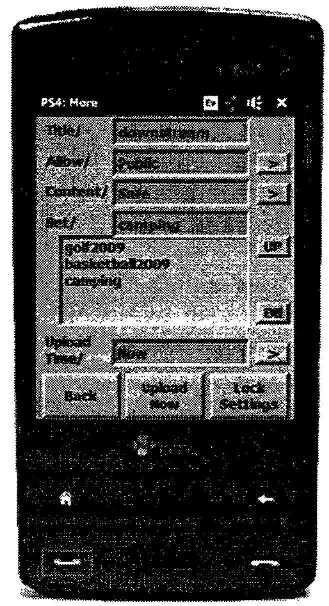


Figure 3.2.6. The More window

3.2.6 The More window

The More window, figure 3.2.6, allows for the specification of four Flickr-defined upload properties relative to a photograph's appearance on the user's account: (1) title, (2) viewing audience, (3) content type and (4) set. These settings may be saved using the [Lock Settings] command, writing the same viewing audience, content type and set values to newly captured photographs. The upload properties are summarized below. Table 3.2.6a defines the More window's voice commands and their graphical counterparts.

1. Title – The photograph's timestamp will be used as its initial title. In figure 3.2.6 the default title has been replaced with "downstream".
2. Viewing audience (Allow) – The viewing audience restricts certain groups from viewing a user's uploaded photographs. Flickr allows for the specification of one audience from the set [Public, Family + Friends, Family, Friends, Just Me].
3. Content type (Content) – The content type defines the photograph's nature. Flickr allows for the specification of one content type from the set [Safe, Moderate, Restricted].

4. Set – A set is a photographic collection. By default a photograph belongs to no specific set; however, in figure 3.2.6 the user has specified "camping" as the upload collection. Our application automatically downloads and presents a user's photosets upon accessing the More window. Selections from this presentation may be made using the stylus or the sets-specific voice commands of table 3.2.6b. Alternatively, new sets may be created by typing a set name in the set text field. A user's photosets are downloaded using the API method of table 3.2.6c.

Graphical Button	Voice Command	Function
Allow >	"Allow"	Scrolls viewing audience
Content >	"Content"	Scrolls content type
UP	"Up" or "Scroll Up"	Scrolls sets in an upward manner
DN	"Down" or "Scroll Down"	Scrolls sets in a downward manner
Upload Time	"Upload Time"	Scrolls upload time
Back	"Back"	Returns to the previous window
Upload Now	"Upload Now"	Opens the Upload window
Lock Settings / Unlock Settings	"Lock Settings" / "Unlock Settings"	Saves current configuration

Table 3.2.6a. Graphical and voice commands of the More window

Voice Command	Function
"*<set>*"	Selects the collection specified by *<set>* for photograph upload

Table 3.2.6b. Sets-specific voice commands of the More window

Method Name	Parameters	XML Response Parsers
flickr.photosets.getList	api_key auth_token user_id api_sig	"<photoset id"> "<title>"

Table 3.2.6c. REST-HTTP GET/POST actions outline for the More window

The More window allows for users to upload photographs at times of most convenience. Upload times may take value from the set [Later, Never, Now, Uploaded]. Newly captured photographs default to the value [Later] and successfully uploaded photographs become marked with value [Uploaded]. All information of the More window is stored in time-of-capture text files (*<timestamp>*_More.txt).

3.2.7 The Upload window

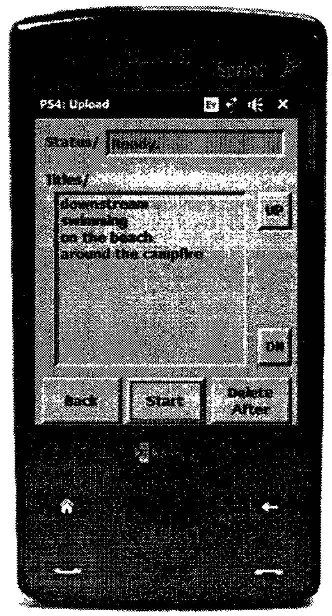


Figure 3.2.7. The Upload window

The Upload window, figure 3.2.7, symbolizes the gateway between the device and the user's Flickr account. Here, users are allowed to make last-minute modifications to all photographs pending upload. Table 3.2.7a defines the Upload window's voice commands and their graphical counterparts. A photograph may be reviewed by selecting the appropriate title graphically or using the titles-specific voice commands of table 3.2.7b.

Graphical Button	Voice Command	Function
UP	"Up" or "Scroll Up"	Scrolls titles in an upward manner

Table 3.2.7a. Graphical and voice commands of the Upload window

Graphical Button	Voice Command	Function
DN	"Down" or "Scroll Down"	Scrolls titles in a downward manner
Back	"Back"	Returns to the previous window
Start	"Start"	Initiates upload
Delete After		Deletes successfully uploaded photographs

Table 3.2.7a (continued). Graphical and voice commands of the Upload window

Voice Command	Function
"<*title*>"	Opens the photograph specified by <*title*> for editing

Table 3.2.7b. Titles-specific voice commands of the Upload window

Like the Tag Bank window, the Upload window also functions in two modes of operation: (1) single photograph upload, entered from the More window as a result of the [Upload Now] command, and (2) multiple photograph upload, entered from the Manager window as a result of the [Upload] command. Users have the option of deleting successfully uploaded photographs from the device after the upload procedure, which relays its progress through the window's status bar. Notice the title of "downstream" which had originated from the More window of figure 3.2.6. Table 3.2.7c outlines the API methods utilized within the Upload window.

Method Name	Parameters	XML Response Parsers
flickr.photosets.create	api_key auth_token primary_photo_id title api_sig	"<photoset id>"

Table 3.2.7c. REST-HTTP GET/POST actions outline for the Upload window

Method Name	Parameters	XML Response Parsers
flickr.photosets.addPhoto	api_key auth_token photoset_id photo_id api_sig	“<rsp stat>” (error code)
flickr.photos.delete	api_key auth_token photo_id api_sig	“<rsp stat”> (error code)

Table 3.2.7c (continued). REST-HTTP GET/POST actions outline for the Upload window

Uploading of the photograph itself requires a multipart/form data REST-HTTP POST to the endpoint URL: http://api.flickr.com/services/upload/. Table 3.2.7d lists the parameters of this action.

Parameters		
api_key	auth_token	content_type
hidden	is_family	is_friend
is_public	safety_level	tags
title	Photo	api_sig

XML Response Parsers
“<photo id>”

Table 3.2.7d. Multipart/form data REST-HTTP GET/POST actions outline for the Upload window

3.3 Development of the data collection server

The P54 server architecture, running communication software complimentary to that of our application, was extended to support the demands of this research. This server provides for an anytime/anywhere update of a participant’s progress logged via cellular connection. Data transmissions of interest are detailed in the diagram of figure 3.3. Files regarding a photograph’s tags and More window content have previously been described. The four remaining files of a transfer packet are of experimental concern and will be

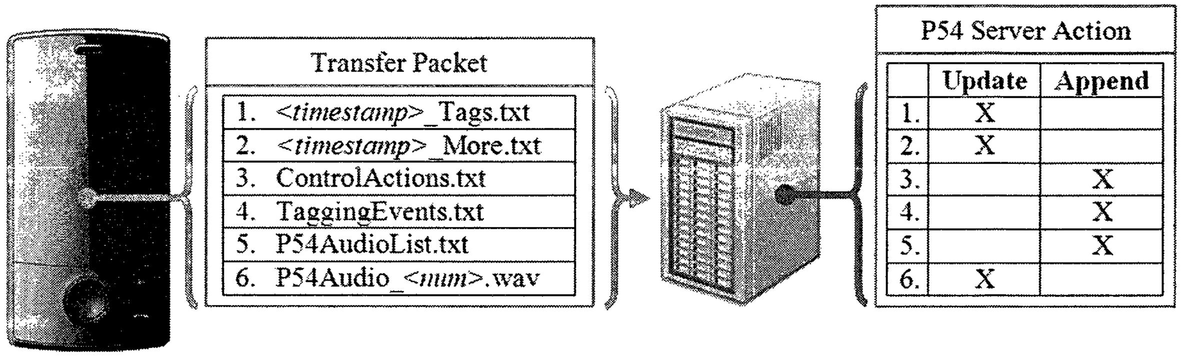


Figure 3.3. Data transmissions and corresponding P54 server actions

discussed in detail in the sections to come; here, we wish to emphasize the server's action (update or append) when presented with files of the various types.

The data within some logs will become invalid. A photograph's tags and More window content, for instance, may change only once throughout a participant's usage of our application; but, when that change occurs it voids all previous data pertaining to the particular photograph. Files of this type require the server to perform an updating operation. Others logs, which focus more on overall application usage, require the server to perform an appending operation. The data existing within these files remains valid and must not be overwritten.

3.4 Experiment with human participants

3.4.1 Participants

To generate interest, students of the College of Engineering and Physical Sciences from the University of New Hampshire were emailed, informing them of the study and noting their usage of a borrowed device if they chose to partake. To best evaluate use of our tagging application the respondents were subdivided into two participant groupings: (1) single-day trials and (2) four-day trials. Sixteen participants took part in our single-day trials and four participants took part in our four-day trials. Care was taken in counterbalancing gender and keyboard type for each participant grouping so that statistical significance testing could be performed properly.

3.4.2 Procedure

The consent and release forms of appendix A and the pre-experiment questionnaire of appendix B were filled out by each participant upon meeting. The initialization procedures of the tag bank required that each participant then create a Flickr account (if at the time they were not a member of the service). Once created participants reviewed two instructional videos [30] entitled "Linking with Flickr" and "Capturing and tagging photographs". The first video demonstrates how the association between the device and a user's Flickr account is made, while the second focuses more on the primary functions of our application. In an effort to further familiarize participants with the application, each was required to complete the on-device tutorial, the prompts of which have been included in appendix C. Before their departure each participant was provided with the written tutorial of appendix D. No emphasis was made on which method of tagging a participant should pursue (GUI or SUI); however, we did emphasize that at least twenty five photographs should be captured/tagged. Participants were compensated with $20 as check or gift card upon their return at which time they were asked to complete the post-experiment questionnaire of appendix B.

3.4.3 Data collection

Data is to be automatically uploaded from a participant's device to the P54 server via cellular connection. Data of interest are: (1) files regarding a photograph's tags and

More window content, (2) logging files, which consist of control action, tagging event and audio recording logs and (3) the audio recordings themselves (we do not see it as being a necessity to collect the participants' photographs since quantitative analyses can be performed through evaluation of the aforementioned data of interest and since many of their photographs may be uploaded to Flickr through usage of our application). The control action logs document the participants' methods of performing some high-level action, which may be invoked either through GUI or SUI. The tagging event logs are more specific to the state of the application and the information being paired with photographs (the tags). For organizational purposes, recorded data will be labeled with the participant's Flickr identification number.

3.4.4 Specifics of the control action logs

Table 3.4.4a lists all control actions of our application. Table 3.4.4b details the logging format of these actions and figure 3.4.4 demonstrates their creation.

Control Action	Active Window	Description
Done	First Use	First Use window exited
	Manager	External camera utility closed
	Tag Bank	Toggled to normal mode
Tutorial	Manager	On-device tutorial opened
Upload	Manager	Upload window opened
Exit	Manager	Tagging application exited
Previous	Manager	Photographs scrolled in an earlier-date manner
	Tags	
Previous_Previous	Manager	Photographs repeatedly scrolled in an earlier-date manner
	Tags	
Rotate	Manager	Photograph rotated counterclockwise
Next	Manager	Photographs scrolled in a latter-date manner
	Tags	

Table 3.4.4a. Control actions, active windows and descriptions

Control Action	Active Window	Description
Next_Next	Manager	Photographs repeatedly scrolled in a latter-date manner
	Tags	
View_Tag_Bank	Manager	Tag Bank window opened
	Tags	
Tags	Manager	Tags window opened
Capture	Manager	External camera utility opened
Delete	Manager	Photograph deleted
Up	Tags	Window's listing scrolled in an upward manner
	Tag Bank	
	More	
	Upload	
Up_Up	Tags	Window's listing repeatedly scrolled in an upward manner
	Tag Bank	
	More	
	Upload	
Down	Tags	Window's listing scrolled in a downward manner
	Tag Bank	
	More	
	Upload	
Down_Down	Tags	Window's listing repeatedly scrolled in a downward manner
	Tag Bank	
	More	
	Upload	
Back	Tags	Returned to the previous window
	Tag Bank	
	More	
	Upload	
More	Tags	More window opened
Alphabetize	Tag Bank	Tag bank reorganized
Sort_By_Uses	Tag Bank	
Edit	Tag Bank	Toggled to editing mode
Next_Allow	More	Viewing audience scrolled
Next_Upload_Time	More	Upload time scrolled
Upload_Now	More	Upload window opened
Next_Content	More	Content type scrolled
Lock_Settings	More	Current configuration saved
Unlock_Settings	More	Default configuration
Sets_Pair	More	Set selected for upload

Table 3.4.4a (continued). Control actions, active windows and descriptions

Control Action	Active Window	Description
Start	Upload	Upload initiated
Delete_After	Upload	Successfully uploaded photographs deleted
Titles_Open	Upload	Tags window opened
Do_Not_Delete_After	Upload	Successfully uploaded photographs not deleted
Insert_Tag	Tags	Newly typed tag paired with photograph
	Tag Bank	
Copy_Tag	Tags	Tag copied from tag bank to photograph
	Tag Bank	
Remove_Tag	Tags	Tag removed from photograph
	Tag Bank	Tag removed from tag bank
Undo	Tags	Lastly added/removed tag removed/reassigned from/to photograph
	Tag Bank	

Table 3.4.4a (continued). Control actions, active windows and descriptions

Timestamp	Active Window	Type	Control Action

Table 3.4.4b. Logging format of control actions

Note: A control action's type may take value from the set [GUI, SUI].

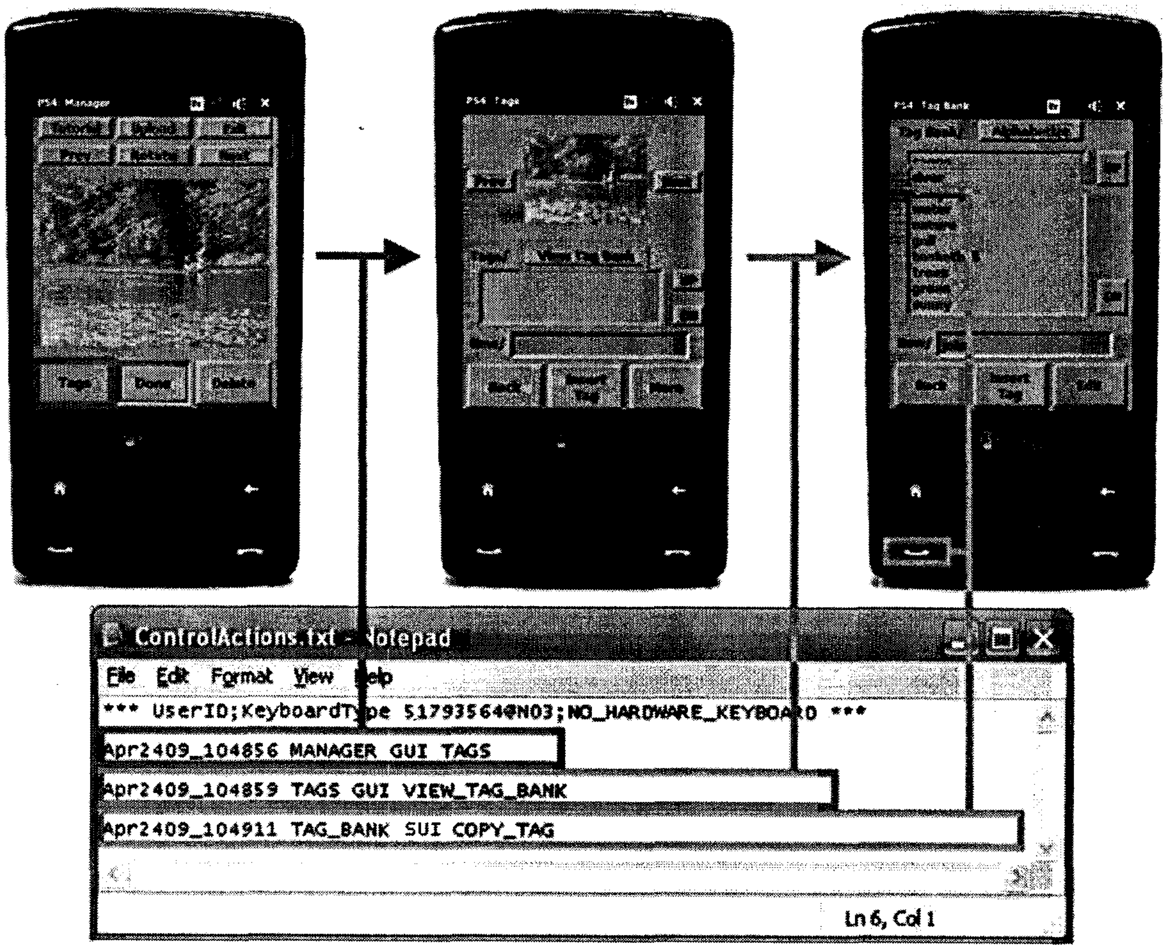


Figure 3.4.4. Creation of control action log data

3.4.5 Specifics of the tagging event logs

Table 3.4.5a lists all tagging events of our application. Table 3.4.5b details the logging format of these events and figure 3.4.5 demonstrates their creation.

Tagging Event	Result	Description
User_Tag	Added_Tag_Bank	A tag from the user's Flickr account has been downloaded and added to the tag bank
	Exists_Tag_Bank	The tag downloaded from the user's Flickr account exists in the tag bank
User_Interest	Added_Tag_Bank	A tag from the Flickr community has been downloaded and added to the tag bank
	Exists_Tag_Bank	The tag downloaded from the Flickr community exists in the tag bank
Scroll	Up	The user has issued an upward scrolling command
	Up_Up	The user has repeatedly issued an upward scrolling command
	Down	The user has issued a downward scrolling command
	Down_Down	The user has repeatedly issued a downward scrolling command
Insert_Tag	Added_Tag_Bank	A newly typed tag has been added to the tag bank
	Exists_Tag_Bank	The newly typed tag exists in the tag bank
	Added_Tags	A newly typed tag has been added to the tags listing
	Exists_Tags	The newly typed tag exists in the tags listing
Copy_Tag	Added_Tags	A tag has been copied from the tag bank to the tags listing
	Exists_Tags	The tag copied from the tag bank exists in the tags listing
Remove_Tag	Removed_Tag_Bank	A tag has been removed from the tag bank
	Removed_Tags	A tag has been removed from the tags listing
Undo	Removed_Tag_Bank	Lastly added tag removed from the tag bank
	Added_Tag_Bank	Lastly removed tag reinserted into the tag bank
	Removed_Tags	Lastly added tag removed from the tags listing
	Added_Tags	Lastly removed tag reinserted into the tags listing

Table 3.4.5a. Tagging events, results and descriptions

Timestamp	Active Window	Photograph	Type	Tagging Event	Result	Tag

Table 3.4.5b. Logging format of tagging events

Note: A tagging event's type may take value from the set [GUI, SUI].

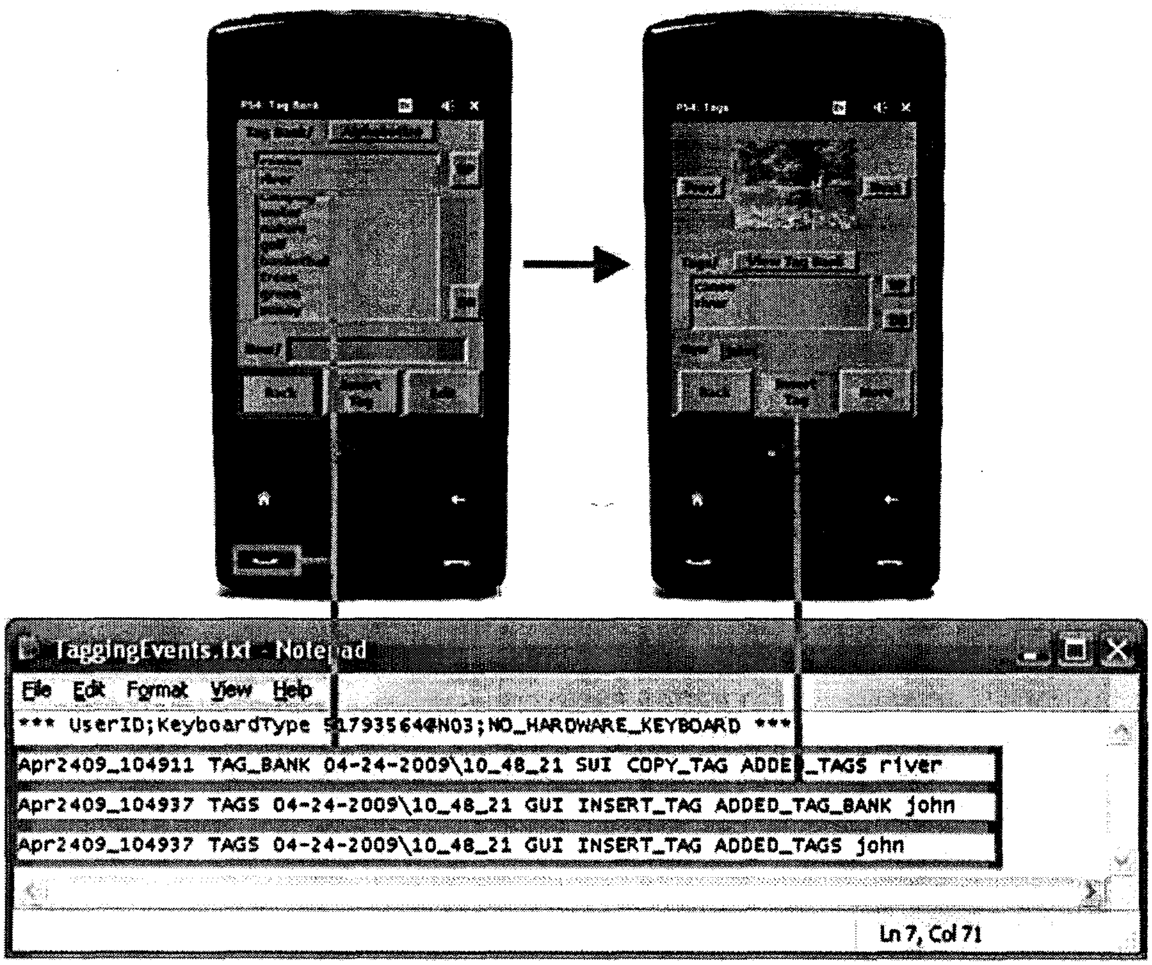


Figure 3.4.5. Creation of tagging event log data

3.4.6 Specifics of the audio recording logs

Table 3.4.6 details the logging format of audio recordings and figure 3.4.6 demonstrates their creation.

File	Active Grammar	Timestamp	Returned Command	Recognized Command

Table 3.4.6. Logging format of audio recordings

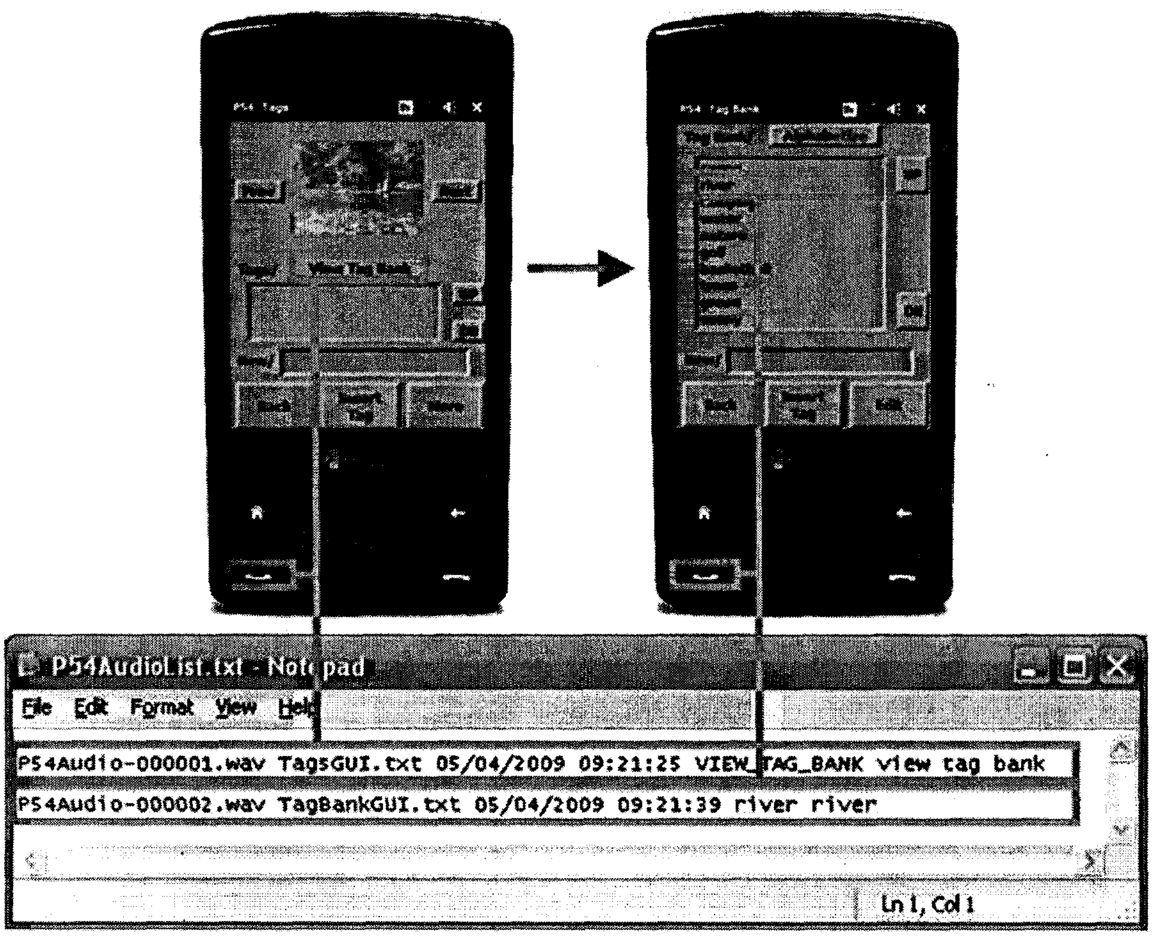


Figure 3.4.6. Creation of audio recording log data

3.5 Data analysis

In the pursuit and investigation of our goals and hypotheses we have proposed a between-subjects experiment with counterbalanced QWERTY/T9 device usage among male and female participants. All pertinent data will be organized by the participants' Flickr identification numbers and logged to the P54 server via cellular connection. Both qualitative and quantitative methods will be used to analyze participants. The qualitative analysis includes pre- and post-experiment Likert scale-based questionnaires concerning the participants' level of familiarity with digital camera usage, their understanding of the concept of tagging photographs and general questioning on the performance of the application and modality preference. The quantitative analysis (outlined below) will categorize the participants' usage of our application based on the information found within their control action, tagging event and audio recording logs. In the sections to follow we present a number of worded mathematical equations. Table 3.5 defines the syntax of these equations.

Syntax	Description	Example
[< *Control Action* >] or [< *Tagging Event* >]	Squared brackets enclose control actions and tagging events	[Removed_Tags] symbolizes all tag removal events upon the tags listing of the Tags window (GUI and SUI commands alike)
[< *Control Action* >].< *property* > or [< *Tagging Event* >].< *property* >	Punctuation is used to access properties (*types* and *results*) of control actions and tagging events	[Removed_Tags].GUI symbolizes GUI-only tag removal events upon the tags listing of the Tags window

Table 3.5. Syntax of analysis equations

3.5.1 Analysis of the control action logs

We foresee a large number of actions from a participant's control action (CA) log to be invoked via GUI and to be scrolling in nature; therefore, to better reveal SUI usages we have chosen to exclude such actions from certain evaluations. We present scrolling actions through summation in equation 3.5.1 below. The sigma operator ($\sum$) simply sums action occurrence numbers from within a participant's control action log, allowing for the construction of many percentage-type equations.

$$\begin{bmatrix}\text{Scrolling}\\ \text{Actions}\end{bmatrix} = \sum \begin{bmatrix}\text{Previous}\\ \text{Previous\_Previous}\\ \text{Next}\\ \text{Next\_Next}\\ \text{Up}\\ \text{Up\_Up}\\ \text{Down}\\ \text{Down\_Down}\\ \text{Next\_Allow}\\ \text{Next\_Content}\\ \text{Next\_Upload\_Time}\end{bmatrix} \qquad \text{Eq. 3.5.1}$$

We represent the scrolling percentage for control actions (equation 3.5.2) as the [Scrolling Actions] to [All Actions] ratio. The [All Actions] grouping symbolizes the total number of control actions from within a participant's control action log.

$$\begin{matrix}\text{CA Scrolling}\\ \text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{Scrolling}\\ \text{Actions}\end{bmatrix}}{\begin{bmatrix}\text{All}\\ \text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.2}$$

The CA GUI non-scrolling percentage (equation 3.5.3) represents control actions issued via GUI excluding those from the [Scrolling Actions] grouping. The property operator (.) extracts GUI-only occurrence numbers from the [All Actions] and [Scrolling Actions] groupings, removing SUI actions from the analysis.

$$\begin{matrix}\text{CA GUI Non-Scrolling} \\ \text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{All} \\ \text{Actions}\end{bmatrix}.\text{GUI} - \begin{bmatrix}\text{Scrolling} \\ \text{Actions}\end{bmatrix}.\text{GUI}}{\begin{bmatrix}\text{All} \\ \text{Actions}\end{bmatrix} - \begin{bmatrix}\text{Scrolling} \\ \text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.3}$$

Similarly, the CA SUI non-scrolling percentage (equation 3.5.4) represents control actions (also excluding those from the [Scrolling Actions] grouping) issued via SUI. The property operator (.) extracts SUI-only occurrence numbers from the [All Actions] and [Scrolling Actions] groupings, removing GUI actions from the analysis.

$$\begin{matrix}\text{CA SUI Non-Scrolling} \\ \text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{All} \\ \text{Actions}\end{bmatrix}.\text{SUI} - \begin{bmatrix}\text{Scrolling} \\ \text{Actions}\end{bmatrix}.\text{SUI}}{\begin{bmatrix}\text{All} \\ \text{Actions}\end{bmatrix} - \begin{bmatrix}\text{Scrolling} \\ \text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.4}$$

Equation 3.5.5, as a note, demonstrates that a participant's CA GUI and SUI non-scrolling percentages may be summed, the result of which represents their modality-independent non-scrolling percentage.

$$\begin{matrix}\text{CA Non-Scrolling} \\ \text{Percentage}\end{matrix} = \begin{matrix}\text{CA GUI Non-Scrolling} \\ \text{Percentage}\end{matrix} + \begin{matrix}\text{CA SUI Non-Scrolling} \\ \text{Percentage}\end{matrix} = 100\ \% \qquad \text{Eq. 3.5.5}$$

Actions created through the selection of information from one of our application's four text listings are of primary interest – it is our belief that the SUI modality will thrive for such interactions. We present selection actions through summation in equation 3.5.6 below. Descriptions of each action of the summation, extracted from table 3.4.4a, then follow in table 3.5.1.

$$\begin{bmatrix}\text{Selection} \\ \text{Actions}\end{bmatrix} = \sum \begin{bmatrix}\text{Copy\_Tag} \\ \text{Remove\_Tag} \\ \text{Sets\_Pair} \\ \text{Titles\_Open}\end{bmatrix} \qquad \text{Eq. 3.5.6}$$

Selection Action	Active Window	Description
Copy_Tag	Tags Tag Bank	Tag copied from tag bank to photograph

Table 3.5.1. Selection actions, active windows and descriptions

Selection Action	Active Window	Description
Remove_Tag	Tags	Tag removed from photograph
	Tag Bank	Tag removed from tag bank
Sets_Pair	More	Set selected for upload
Titles_Open	Upload	Tags window opened

Table 3.5.1 (continued). Selection actions, active windows and descriptions

We represent the SUI selections percentage (equation 3.5.7) as the [Selection Actions].SUI to [Selection Actions] ratio. GUI actions have been removed from this analysis through the property operator (.).

$$\begin{matrix}\text{SUI Selections} \\ \text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{Selection} \\ \text{Actions}\end{bmatrix}.\text{SUI}}{\begin{bmatrix}\text{Selection} \\ \text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.7}$$

3.5.2 Analysis of the tagging event logs

As in our analysis of the control action logs, some events of the tagging event (TE) logs may also be omitted. The events of [User_Tag] and [User_Interest] are created, with no contribution by a participant, by the Flickr API during the tag bank initialization procedure. To better reveal usages we have chosen to exclude such Flickr events, presented through summation in equation 3.5.8 below, from certain evaluations.

$$\begin{bmatrix}\text{Flickr} \\ \text{Events}\end{bmatrix} = \sum \begin{bmatrix}\text{User\_Tag} \\ \text{User\_Interest}\end{bmatrix} \qquad \text{Eq. 3.5.8}$$

We represent the scrolling percentage for tagging events (equation 3.5.9) as the [Scroll] to [All Events] ratio excluding [Flickr Events]. The [All Events] grouping symbolizes the total number of tagging events found within a participant's tagging event log.

$$\begin{matrix}\text{TE Scrolling}\\ \text{Percentage}\end{matrix} = 100\frac{[\text{Scroll}]}{\begin{bmatrix}\text{All}\\ \text{Events}\end{bmatrix} - \begin{bmatrix}\text{Flickr}\\ \text{Events}\end{bmatrix}} \qquad \text{Eq. 3.5.9}$$

The TE GUI non-scrolling percentage (equation 3.5.10) represents tagging events issued via GUI excluding those of [Scroll] and from the [Flickr Events] grouping. The property operator (.) extracts GUI-only occurrence numbers from the [Scroll] event and [All Events] and [Flickr Events] groupings, removing SUI events from the analysis.

$$\begin{matrix}\text{TE GUI Non-Scrolling}\\ \text{Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{All}\\ \text{Events}\end{bmatrix}.\text{GUI} - \begin{bmatrix}\text{Flickr}\\ \text{Events}\end{bmatrix} - [\text{Scroll}].\text{GUI}}{\begin{bmatrix}\text{All}\\ \text{Events}\end{bmatrix} - \begin{bmatrix}\text{Flickr}\\ \text{Events}\end{bmatrix} - [\text{Scroll}]} \qquad \text{Eq. 3.5.10}$$

The TE SUI non-scrolling percentage (equation 3.5.11) represents tagging events (also excluding those of [Scroll] and from the [Flickr Events] grouping) issued via SUI. The property operator (.) extracts SUI-only occurrence numbers from the [Scroll] event and [All Events] and [Flickr Events] groupings, removing GUI events from the analysis.

$$\begin{matrix}\text{TE SUI Non-Scrolling}\\ \text{Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{All}\\ \text{Events}\end{bmatrix}.\text{SUI} - \begin{bmatrix}\text{Flickr}\\ \text{Events}\end{bmatrix} - [\text{Scroll}].\text{SUI}}{\begin{bmatrix}\text{All}\\ \text{Events}\end{bmatrix} - \begin{bmatrix}\text{Flickr}\\ \text{Events}\end{bmatrix} - [\text{Scroll}]} \qquad \text{Eq. 3.5.11}$$

Equation 3.5.12, as a note, demonstrates that a participant's TE GUI and SUI non-scrolling percentages for tagging events may be summed, the result of which represents their modality-independent non-scrolling percentage.

$$\begin{matrix}\text{TE Non-Scrolling}\\ \text{Percentage}\end{matrix} = \begin{matrix}\text{TE GUI Non-Scrolling}\\ \text{Percentage}\end{matrix} + \begin{matrix}\text{TE SUI Non-Scrolling}\\ \text{Percentage}\end{matrix} = 100\ \% \qquad \text{Eq. 3.5.12}$$

As previously stated, actions created through the selection of information from one of our application's four text listings (first evaluated by the selections percentage for control actions) are of primary interest. The annotating events, presented through summation in equation 3.5.13 below, of [Insert_Tag], [Copy_Tag], [Remove_Tag] and

[Undo] are specific to the adding and removing of tags to and from the tags and tag bank listings. An overlap between [Annotating Events] and [Selection Actions] does exist, namely in the [Copy_Tag] and [Remove_Tag] events/actions; yet, it is important to remember that the two sets provide for distinct analyses. An analysis upon control actions (the [Selection Actions] grouping) would reveal only general GUI/SUI statistics, whereas an analysis upon tagging events (the [Annotating Events] grouping) would reveal more detailed information regarding the state of our application. Annotating events consider not only the contents of the tags and tag bank listings, but also how their entries were generated (either through GUI, SUI or keyboard). They deepen the general usage statistics offered by the selections percentage, and may answer questions like: why (or why not) was the SUI chosen for some tagging interaction? A description of each annotating event, extracted from table 3.4.5a, follows equation 3.5.13 in table 3.5.2.

$$\begin{bmatrix} \text{Annotating} \\ \text{Events} \end{bmatrix} = \sum \begin{bmatrix} \text{Insert\_Tag} \\ \text{Copy\_Tag} \\ \text{Remove\_Tag} \\ \text{Undo} \end{bmatrix} \qquad \text{Eq. 3.5.13}$$

Annotating Event	Result	Description
Insert_Tag	Added_Tag_Bank	A newly typed tag has been added to the tag bank
	Exists_Tag_Bank	The newly typed tag exists in the tag bank
	Added_Tags	A newly typed tag has been added to the tags listing
	Exists_Tags	The newly typed tag exists in the tags listing
Copy_Tag	Added_Tags	A tag has been copied from the tag bank to the tags listing
	Exists_Tags	The tag copied from the tag bank exists in the tags listing
Remove_Tag	Removed_Tag_Bank	A tag has been removed from the tag bank
	Removed_Tags	A tag has been removed from the tags listing

Table 3.5.2. Annotating events, results and descriptions

Annotating Event	Result	Description
Undo	Removed_Tag_Bank	The lastly added tag has been removed from the tag bank
	Added_Tag_Bank	The lastly removed tag has been reinserted into the tag bank
	Removed_Tags	The lastly added tag has been removed from the tags listing
	Added_Tags	The lastly removed tag has been reinserted into the tags listing

Table 3.5.2 (continued). Annotating events, results and descriptions

We foresee a large number of events from the [Annotating Events] grouping to be events of inserting newly typed tags. The tag insertion percentage (TIP), of equation 3.5.14 represents this [Insert_Tag] to [Annotating Events] ratio.

$$\begin{matrix}\text{Tag Insertion}\\ \text{Percentage}\end{matrix}(\text{TIP}) = 100\,\frac{[\text{Insert\_Tag}]}{\left[\begin{matrix}\text{Annotating}\\ \text{Events}\end{matrix}\right]} \qquad \text{Eq. 3.5.14}$$

The tag duplication percentage (equation 3.5.15) will reveal a participant's awareness of the tag bank's contents. We base this ratio solely on [Insert_Tag] events since they result from newly typed tag insertions. The property operator (.) checks for the tag's existence within the tag bank, thereby informing us on the participant's awareness of its contents. The tag duplication and tag insertion percentages are GUI-only percentages.

$$\begin{matrix}\text{Tag Duplication}\\ \text{Percentage}\end{matrix} = 100\,\frac{[\text{Insert\_Tag}].\text{Exists\_Tag\_Bank}}{[\text{Insert\_Tag}]} \qquad \text{Eq. 3.5.15}$$

We use the tag bank utilization percentage (equation 3.5.16) to distinguish between tags paired with photographs via keyboard and those copied from the tag bank. The property operator (.) excludes [Insert_Tag] events issued with the Tag Bank window in its editing mode. These tags would not be paired with photographs and must be omitted from the utilization ratio.

$$\begin{matrix}\text{Tag Bank} \\ \text{Utilization Percentage}\end{matrix} = 100\frac{[\text{Copy\_Tag}]}{[\text{Copy\_Tag}] + [\text{Insert\_Tag}] - [\text{Insert\_Tag}].\text{Tag\_Bank\_Edit}} \qquad \text{Eq. 3.5.16}$$

The GUI and SUI non-TIP tagging percentages (equations 3.5.17 and 3.5.18) form our final classification on modality preference. Each excludes the [Insert_Tag] event (keyboard input) from the [Annotating Events] grouping and utilizes the property operator (.GUI/.SUI) to extract GUI and SUI [Copy_Tag], [Remove_Tag] and [Undo] events.

$$\begin{matrix}\text{GUI Non-TIP} \\ \text{Tagging Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix}.\text{GUI} - [\text{Insert\_Tag}]}{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix} - [\text{Insert\_Tag}]} \qquad \text{Eq. 3.5.17}$$

$$\begin{matrix}\text{SUI Non-TIP} \\ \text{Tagging Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix}.\text{SUI}}{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix} - [\text{Insert\_Tag}]} \qquad \text{Eq. 3.5.18}$$

As a note, equation 3.5.19 demonstrates that a participant's GUI and SUI non-TIP tagging percentages may be summed, the result of which represents their modality-independent non-TIP tagging percentage.

$$\begin{matrix}\text{Non-TIP} \\ \text{Tagging Percentage}\end{matrix} = \begin{matrix}\text{GUI Non-TIP} \\ \text{TaggingPercentage}\end{matrix} + \begin{matrix}\text{SUI Non-TIP} \\ \text{TaggingPercentage}\end{matrix} = 100\ \% \qquad \text{Eq. 3.5.19}$$

Figure 3.5.2 depicts the annotating events sample space and its partitions. As shown in the upper-right corner of the figure, the outermost oval represents all events from the [Annotating Events] grouping. Each inner oval segregates some number of these events from the whole, corresponding by color to the numerators of our tagging event logs analysis equations. For example, the numerator of the SUI non-TIP tagging percentage, encircled in purple in figure 3.5.2, consists of the SUI events [Copy_Tag], [Remove_Tag] and [Undo]. Each of these events has a certain list of possible results listed directly below the event. It is the numerical counts of these results from within a participant's tagging event log which construct the numerator of the SUI non-TIP tagging percentage. Note that the only equation to span between GUI/SUI modalities is the tag bank utilization percentage.

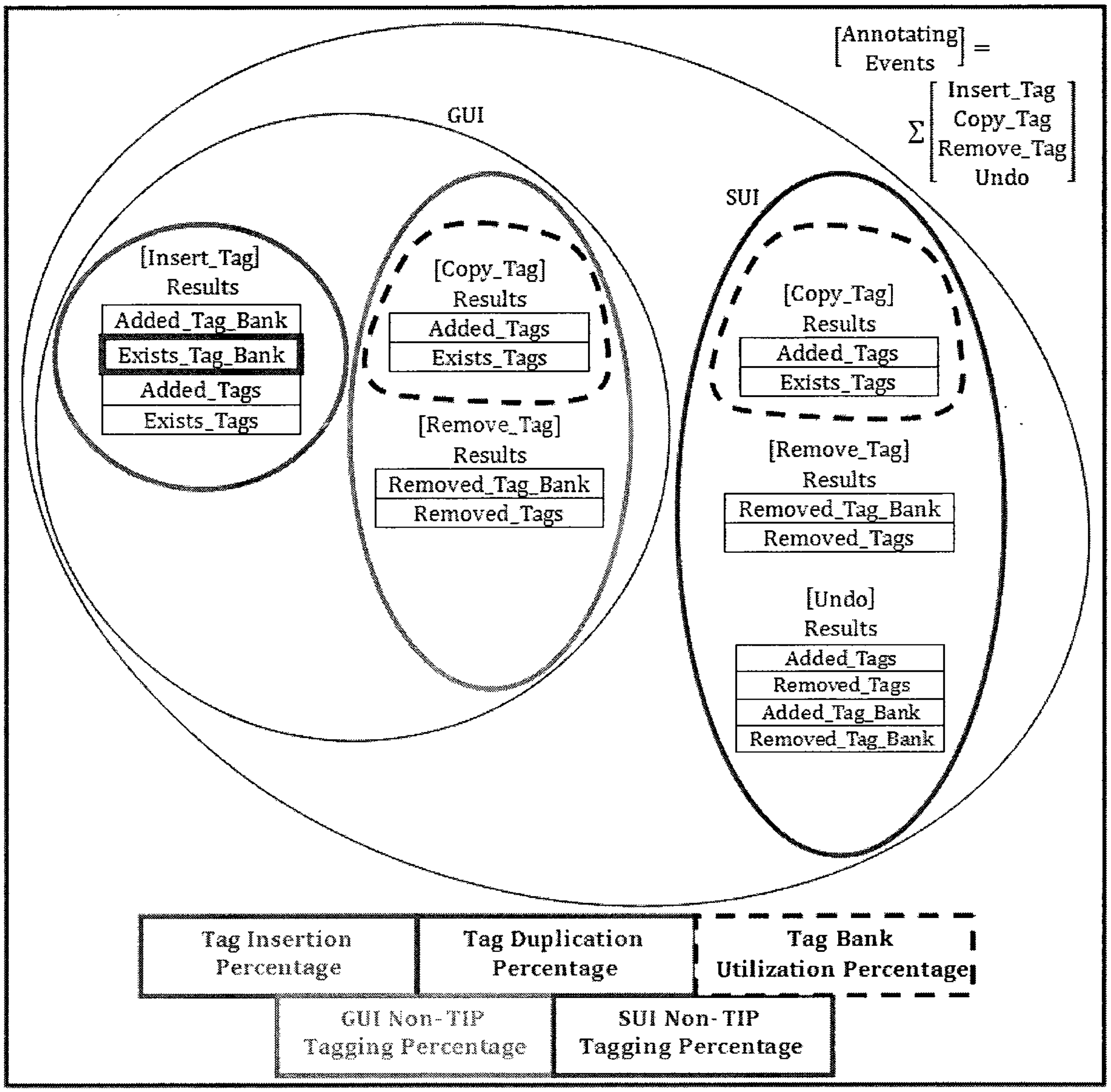


Figure 3.5.2. The annotating events sample space and its partitions

3.5.3 Analysis of the audio recording logs

Table 3.5.3, inspired through the works of Paek et al. [31], detail the top-down classifications to be used in the evaluation of a participant's audio recordings.

Classification	**Description**	**Incorrect Saying Example** **(Correct Saying: "remove tag river bed")**
Correct Recognition	Properly recognized utterance	N/A
Out-of-Grammar	Grammar does not support utterance	N/A
Substitution	Additional wording within utterance	"remove the tag river bed"
Segmentation	Incomplete utterance	"remove tag river"
Deletion	Necessary words removed from utterance	"remove river bed"
Order Rearrangement	Incorrect ordering within utterance	"river bed remove tag"
Disfluency	Poor articulation within utterance	"remove tag riv bed"
Noisy Environment	Noise overtakes utterance	N/A
Unrecognized	Utterance supported by grammar yet unrecognized	N/A

Table 3.5.3. Classifications of audio recordings, their descriptions and saying examples

Equations 3.5.21 and 3.5.22 describe a participant's level of success in using the SUI. Recognition rate (equation 3.5.21), an inflated version of task completion rate (equation 3.5.22), removes seven of nine classifications from the [All Classifications] grouping (equation 3.5.20) in an attempt to account for user error. Recognition rate is a sort of "could have been/would have been" performance measure, while task completion rate is a performance measure of true participant experiences. The [All Classifications] grouping represents a participant's total number of utterances.

$$\begin{bmatrix} \text{All} \\ \text{Classifications} \end{bmatrix} = \sum \begin{bmatrix} \text{Correct Recognition} \\ \text{Out-of-Grammar} \\ \text{Substitution} \\ \text{Segmentation} \\ \text{Deletion} \\ \text{Order Rearrangement} \\ \text{Disfluency} \\ \text{Noisy Environment} \\ \text{Unrecognized} \end{bmatrix} \qquad \text{Eq. 3.5.20}$$

$$\begin{matrix} \text{Recognition} \\ \text{Rate} \end{matrix} = 100 \frac{[\text{Correct Recognitions}]}{[\text{Correct Recognitions}] + [\text{Unrecognized}]} \qquad \text{Eq. 3.5.21}$$

$$\begin{matrix} \text{Task} \\ \text{Completion Rate} \end{matrix} = 100 \frac{[\text{Correct Recognitions}]}{[\text{All Classifications}]} \qquad \text{Eq. 3.5.22}$$

3.5.4 Testing for statistical significance

Determining statistical significance between two variables will be achieved using the one-way repeated-measures ANOVA assessment. Dependent and independent variables are as follows:

Variable Type	Control Actions	Tagging Events
Dependent	CA Scrolling Percentage CA GUI Non-Scrolling Percentage CA SUI Non-Scrolling Percentage SUI Selections Percentage Number of Captured Photographs Task Completion Rate	TE Scrolling Percentage TE GUI Non-Scrolling Percentage TE SUI Non-Scrolling Percentage Number of Tags Generated Tag Insertion Percentage Tag Duplication Percentage Tag Bank Utilization Percentage GUI Non-TIP Tagging Percentage SUI Non-TIP Tagging Percentage Task Completion Rate
Independent	Keyboard Type Gender Control Action Level Task Completion Level	Keyboard Type Gender Tagging Event Level Task Completion Level

Table 3.5.4. Listing of statistical significance testing variables

CHAPTER 4

EVALUATION AND DISCUSSION

4.1 The corpus

4.1.1 Summary

Considered as one of this works greatest achievements, the gathering of the corpus not only allowed for the present evaluation of our application but also provides an initial dataset for future evaluations. Table 4.1.1 outlines the corpus and figures 4.1.1a through 4.1.1d depict a collection of sample photographs and their respective tags.

	Single Day Participants	Four Day Participants	Total
Photographs	436	141	577
Tags Files	436	141	577
More Properties Files	436	141	577
Control Action Logs	16	4	20
Tagging Event Logs	16	4	20
Tag Bank Files	16	4	20
Pre-experiment Questionnaires	16	4	20
Post-experiment Questionnaires	16	4	20
Audio Recordings/Logs	1036	68	1104
Total	2424	511	2935

Table 4.1.1. Corpus outline

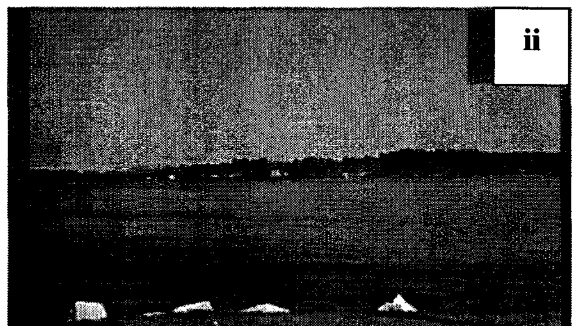
In quick review of the figures below we gain some insight to the varying photographic environments in which our application was used. The photograph of figure

4.1.1a seemed to have been captured during an ongoing lecture and tagged via GUI as to not disrupt students nearby the photographer. Notice that the photograph possesses a number of tags all seeming to relate to a typical campus environment.

Tagged with:
class room; cluster; girls; shirt; table; water bottle

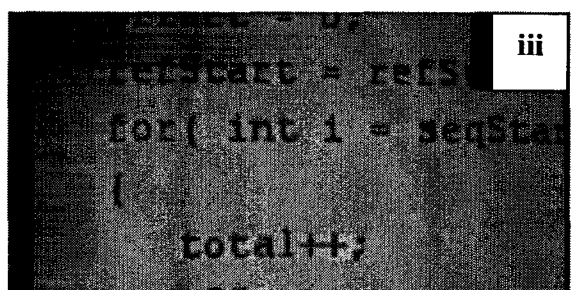
Figure 4.1.1a. Sample photograph (i)

Figures 4.1.1b and 4.1.1c are quite different in setting from that of figure 4.1.1a. Figure 4.1.1b, although more pleasing to the eye, contains significantly fewer tags. Our hypothesis of the SUI being the more convenient tagging modality, if true, should have increased the number of tags for photographs taken in such a setting. Clearly, factors such as environmental noise, out-of-grammar utterances and, perhaps, the participant's general ambition, influence their tagging extent.



Tagged with:
the bay

Figure 4.1.1b. Sample photograph (ii)



Tagged with:
computer; for loops; java; programming; technology

Figure 4.1.1c. Sample photograph (iii)

Interestingly, the photographs of figure 4.1.1d were captured by a participant who had chosen to reuse certain tags among

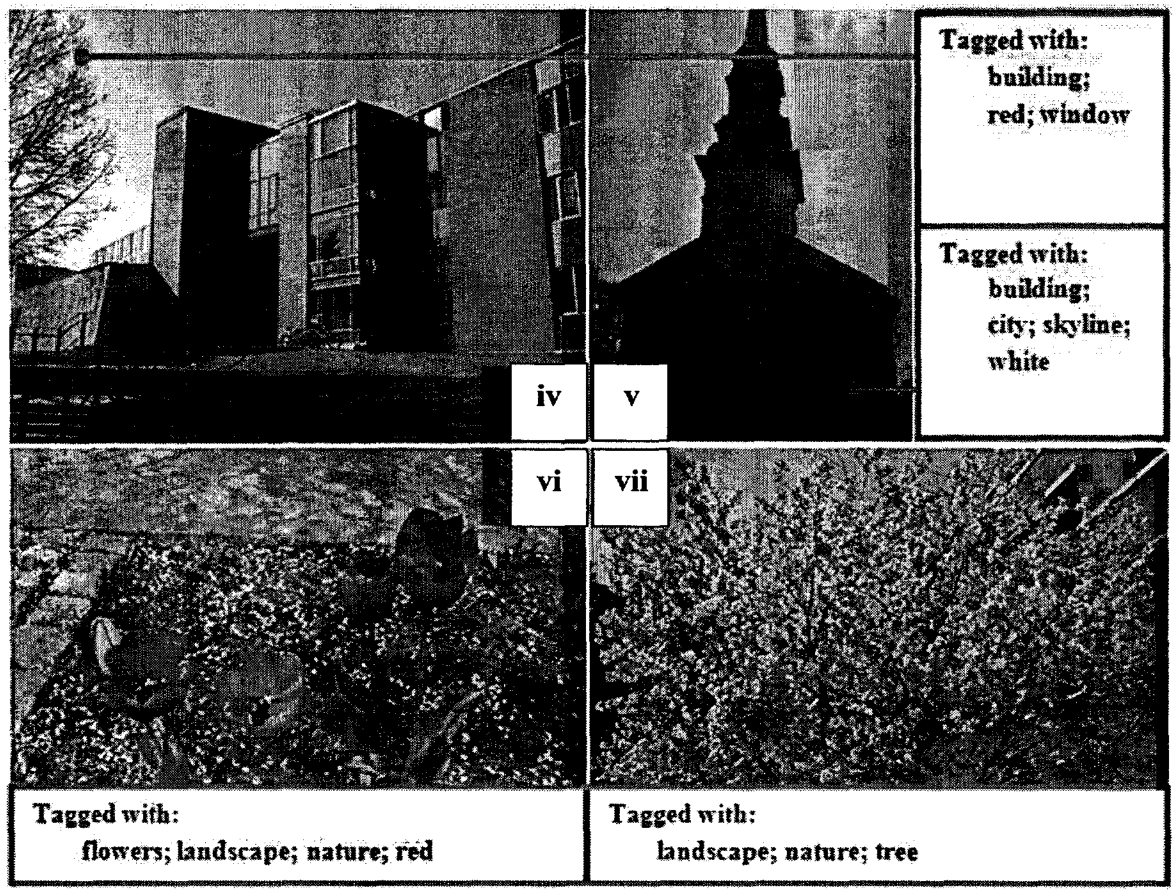


Figure 4.1.1d. Sample photographs (iv-vii)

similar objects. Did our participant understand the concept of the tag bank? What was their modality preference? Is it possible that task completion rate influenced this choice in modality? We next evaluate.

4.1.2 Presentation: significance testing and charting techniques

As previously stated, determining statistical significance between two variables was achieved using the one-way repeated-measures ANOVA assessment (significance between variables when $p < 0.05$). Tests with $p < 0.15$, although not statistically significant, were named as trends. Data is presented graphically, through filtered charts, where we group similar participants by transforming their numeric result of an analysis equation into color-coded worded approximations. For example, the numeric counts of control action totals for participants with totals below the median would be approximated as having [Low] control action totals. The remaining participants, those above the median, would be approximated as having [High] control action totals. In the formation of each chart these approximations are color coded, where we label participants approximated as [Low] with blue markers and those approximated as [High] with red markers. We represent the overall numeric means of our analysis equations with black markers. Such approximations support the graphical representation of a largely varying set (which is the case in many of our evaluations to come). In the title of each chart the phrase "filtered by" is abbreviated as "Fb." (e.g. "Fb. Control Action Totals"). Our evaluations to follow consider single-day participants (SDP) only, from which we attempt to reveal underlying trends to usage as exposure to our application increased (hence our method of filtering by totals). Additional filtering methods will be evaluated in section 4.4. Participants of the four-day type demonstrated similarities in usage to those of the single-day type; but, because the four-day trials consisted of a fewer number of participants, we omit their evaluation from this work.

4.2 Evaluation and discussion of the control action logs

Table 4.2a lists the participants' control action totals under our low/high method of filtering. For convenience, we then restate the analytical equations of the control action logs, the values of which are depicted in figure 4.2 and table 4.2b. As assumed, a large number of participants' actions were invoked via GUI and were scrolling in nature. Therefore, as stated in equations 3.5.3 and 3.5.4, we present the CA GUI and SUI non-scrolling percentages excluding this factor. The SUI selections percentage (SUI actions created through the selection of information from one of our application's four text listings – equation 3.5.7) is also presented in figure 4.2 and table 4.2b.

Filter	Control Action Totals							
Low	183	184	258	298	318	360	456	480
High	522	534	621	646	812	917	1044	1159

Table 4.2a. Control action totals under our low/high method of filtering (single-day participants)

$$\begin{bmatrix}\text{Scrolling}\\\text{Actions}\end{bmatrix} = \sum \begin{bmatrix}\text{Previous}\\\text{Previous\_Previous}\\\text{Next}\\\text{Next\_Next}\\\text{Up}\\\text{Up\_Up}\\\text{Down}\\\text{Down\_Down}\\\text{Next\_Allow}\\\text{Next\_Content}\\\text{Next\_Upload\_Time}\end{bmatrix} \qquad \text{Eq. 3.5.1}$$

$$\begin{matrix}\text{CA Scrolling}\\\text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{Scrolling}\\\text{Actions}\end{bmatrix}}{\begin{bmatrix}\text{All}\\\text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.2}$$

$$\begin{matrix}\text{CA GUI Non-Scrolling}\\\text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{All}\\\text{Actions}\end{bmatrix}.\text{GUI} - \begin{bmatrix}\text{Scrolling}\\\text{Actions}\end{bmatrix}.\text{GUI}}{\begin{bmatrix}\text{All}\\\text{Actions}\end{bmatrix} - \begin{bmatrix}\text{Scrolling}\\\text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.3}$$

$$\begin{matrix}\text{CA SUI Non-Scrolling}\\\text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{All}\\\text{Actions}\end{bmatrix}.\text{SUI} - \begin{bmatrix}\text{Scrolling}\\\text{Actions}\end{bmatrix}.\text{SUI}}{\begin{bmatrix}\text{All}\\\text{Actions}\end{bmatrix} - \begin{bmatrix}\text{Scrolling}\\\text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.4}$$

$$\begin{matrix}\text{SUI Selections}\\\text{Percentage}\end{matrix} = 100\,\frac{\begin{bmatrix}\text{Selection}\\\text{Actions}\end{bmatrix}.\text{SUI}}{\begin{bmatrix}\text{Selection}\\\text{Actions}\end{bmatrix}} \qquad \text{Eq. 3.5.7}$$

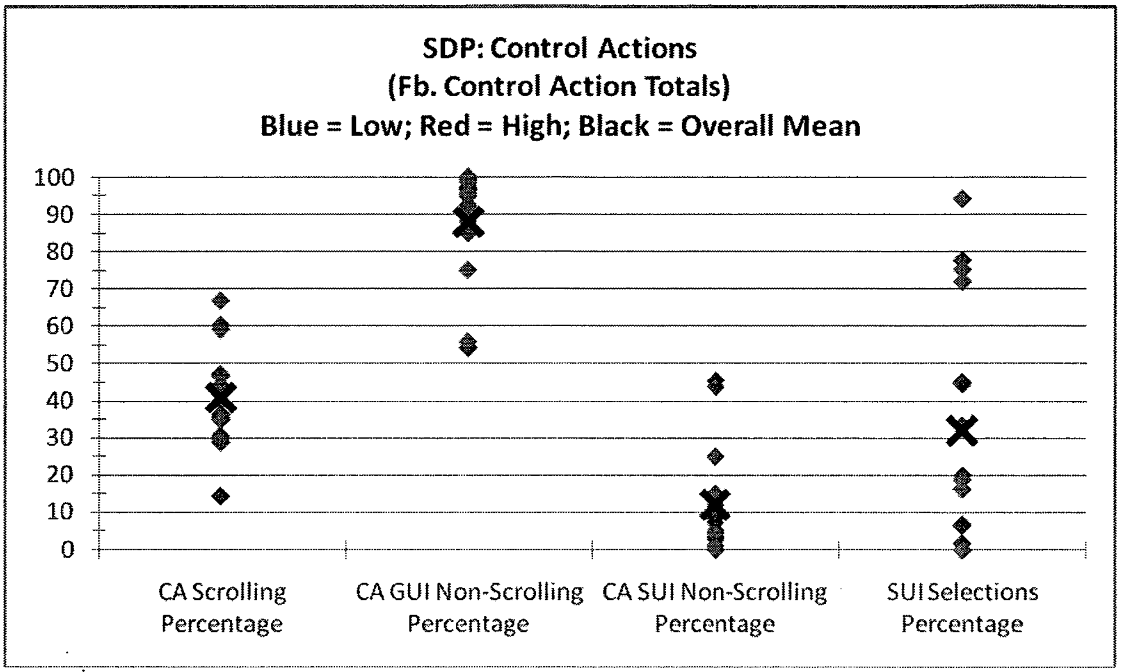


Figure 4.2. Characteristics of control actions filtered by control action totals (single-day participants)

	Filter	CA Scrolling Percentage	CA GUI Non-Scrolling Percentage	CA SUI Non-Scrolling Percentage	SUI Selections Percentage
Mean	**Low**	32.1	90.5	9.6	25.2
	High	49.4	85.6	14.4	38.7
	Overall	40.8	88.0	12.0	32.0
Standard Deviation	**Low**	8.5	14.8	14.8	27.9
	High	11.8	14.5	14.5	36.8
	Overall	13.3	14.4	14.4	32.3

Table 4.2b. Characteristics of control actions filtered by control action totals (single-day participants)

Filtering participants by control action totals reveals their tendencies to invoke scrolling actions as usage of our application increased. Table 4.2c shows that differences between CA scrolling percentages of participants with low and high control action totals are statistically significant ($p < 0.05$), but that differences in the number of photographs captured by these participants (listed in table 4.2d) are not statistically significant

($p \gg 0.15 \gg 0.05$). In combination, these findings imply that participants who ranked highest in control action totals explored our application to a greater extent, not by capturing more photographs, but through the review of previously captured ones and in their scrolling of our application's text listings (the actions of equation 3.5.1). Throughout the remainder of our evaluations we shall refer to such participants (participants with highest totals) as having an increased ambition towards our application.

Independent Variable	**Dependent Variables**	**Mean Difference**	**Significance**
Keyboard Type (QWERTY vs. T9)	**CA Scrolling Percentage**	-15.1	0.02
	CA GUI Non-Scrolling Percentage	-1.5	0.85
	CA SUI Non-Scrolling Percentage	1.5	0.85
	SUI Selections Percentage	14.0	0.40
	Number of Captured Photographs	0.6	0.85
	Task Completion Rate	9.1	0.51
Gender (Male vs. Female)	**CA Scrolling Percentage**	-6.9	0.32
	CA GUI Non-Scrolling Percentage	-0.3	0.97
	CA SUI Non-Scrolling Percentage	0.3	0.97
	SUI Selections Percentage	-17.4	0.30
	Number of Captured Photographs	0.6	0.85
	Task Completion Rate	7.3	0.58
Control Action Level (Low vs. High)	**CA Scrolling Percentage**	-17.2	0.01
	CA GUI Non-Scrolling Percentage	4.8	0.52
	CA SUI Non-Scrolling Percentage	-4.8	0.52
	SUI Selections Percentage	-13.5	0.42
	Number of Captured Photographs	-2.4	0.48
	Task Completion Rate	-14.3	0.28

Table 4.2c. Results of control action statistical significance testing (single-day participants)

Note: Other factors of (or approaching) statistical significance are left to be expanded upon under a future work; specifically, those concerning the independent variables of keyboard type and gender. From table 4.2c, though, it appears that no significant *differences in usage exist, with the exception of the keyboard type-CA scrolling percentage assessment. In this section, and the remaining sections of the chapter, we intend to extract underlying trends to usage as exposure to our application increased, discussing the implications of low/high filtering only.*

Filter	Number of Capture Photographs							
Low	14	19	30	22	32	37	28	14
High	25	34	22	22	33	36	29	25

Table 4.2d. Number of captured photographs under our low/high method of filtering (single-day participants)

Regardless of ambition, and excluding some special cases of increased or nearly equivalent SUI usage, an overwhelming preference towards the GUI can be seen from the declining percentages between the CA GUI and SUI non-scrolling percentages of figure 4.2 (CA GUI non-scrolling percentage mean: 88.0 %, CA SUI non-scrolling percentage mean: 12.0 %). However, when we concentrate on selection actions only, the SUI selections percentage reveals a dwindling GUI preference (best represented by the overall means of figure 4.2, showing CA SUI non-scrolling percentage mean: 12.0 %, SUI selection percentage mean: 32.0 %). We expect this same increase in the percentage of SUI usage to also unfold in the evaluation of the tagging event logs for two reasons: (1) our hypothesis of the SUI being the more convenient tagging modality and (2) manipulations upon the tags and tag bank listings are selection-centered; that is, a properly initialized tag bank should avoid text entry and make itself available for GUI/SUI tag selections.

4.3 Evaluation and discussion of the tagging event logs

4.3.1 First set of analytical equations

In table 4.3.1a we list the participants' tagging event totals under our low/high method of filtering. We then restate below the first set of analytical equations of the tagging event logs, the values of which are depicted in figure 4.3.1 and table 4.3.1b. As with control actions, a large number of participants' tagging events were invoked via GUI and, again, were scrolling in nature. As stated in equations 3.5.10 and 3.5.11, we present the TE GUI and SUI non-scrolling percentages excluding this factor.

Filter	Tagging Event Totals							
Low	31	38	93	104	121	134	139	170
High	180	214	226	231	273	590	733	845

Table 4.3.1a. Tagging event totals under our low/high method of filtering (single-day participants)

$$\begin{bmatrix}\text{Flickr}\\\text{Events}\end{bmatrix} = \sum \begin{bmatrix}\text{User\_Tag}\\\text{User\_Interest}\end{bmatrix} \qquad \text{Eq. 3.5.8}$$

$$\begin{matrix}\text{TE Scrolling}\\\text{Percentage}\end{matrix} = 100\frac{[\text{Scroll}]}{\begin{bmatrix}\text{All}\\\text{Events}\end{bmatrix} - \begin{bmatrix}\text{Flickr}\\\text{Events}\end{bmatrix}} \qquad \text{Eq. 3.5.9}$$

$$\begin{matrix}\text{TE GUI Non-Scrolling}\\\text{Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{All}\\\text{Events}\end{bmatrix}.\text{GUI} - \begin{bmatrix}\text{Flickr}\\\text{Events}\end{bmatrix} - [\text{Scroll}].\text{GUI}}{\begin{bmatrix}\text{All}\\\text{Events}\end{bmatrix} - \begin{bmatrix}\text{Flickr}\\\text{Events}\end{bmatrix} - [\text{Scroll}]} \qquad \text{Eq. 3.5.10}$$

$$\begin{matrix}\text{TE SUI Non-Scrolling}\\\text{Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{All}\\\text{Events}\end{bmatrix}.\text{SUI} - \begin{bmatrix}\text{Flickr}\\\text{Events}\end{bmatrix} - [\text{Scroll}].\text{SUI}}{\begin{bmatrix}\text{All}\\\text{Events}\end{bmatrix} - \begin{bmatrix}\text{Flickr}\\\text{Events}\end{bmatrix} - [\text{Scroll}]} \qquad \text{Eq. 3.5.11}$$

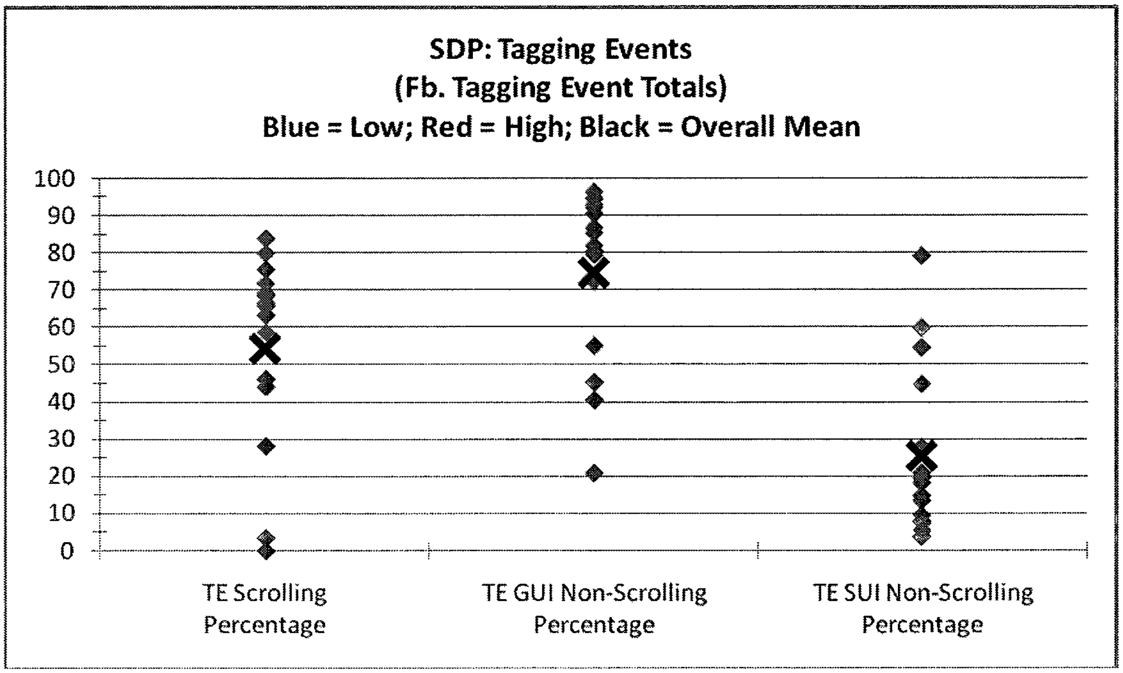


Figure 4.3.1. Characteristics of tagging events filtered by tagging event totals

(first set of analytical equations, single-day participants)

	Filter	TE Scrolling Percentage	TE GUI Non-Scrolling Percentage	TE SUI Non-Scrolling Percentage
Mean	**Low**	39.0	81.4	18.6
	High	69.7	67.8	32.2
	Overall	54.4	74.6	25.4
Standard Deviation	**Low**	27.5	11.6	11.6
	High	8.5	28.7	28.7
	Overall	25.2	22.3	22.3

Table 4.3.1b. Characteristics of tagging events filtered by tagging event totals

(first set of analytical equations, single-day participants)

Similar to our findings in evaluating the control action logs, filtering participants by tagging event totals reveals their tendencies to invoke scrolling events as usage of our application increased. Table 4.3.1c shows that differences between TE scrolling percentages of participants with low and high tagging event totals are again statistically significant with $p < 0.05$. However, contrary to our control action logs evaluation, where we had found an independence to exist between the variables [Number of Captured Photographs] and [Control Action Level], table 4.3.1c shows the increase in the number of tags generated by our more ambitious participants (listed in table 4.3.1d) to be a trend ($p = 0.14$). This significant finding will be expanded upon shortly, but before leaving this portion of the evaluation we note the less overwhelming preference towards the GUI in comparison to that of figure 4.2 for control actions (figure 4.3.1: TE GUI-/SUI-non scrolling percentage means of 74.6 %/25.4 %, figure 4.2: CA GUI-/SUI-non scrolling percentage means of 88.0 %/12.0 %).

Independent Variable	Dependent Variables	Mean Difference	Significance
Keyboard Type (QWERTY vs. T9)	TE Scrolling Percentage	-28.2	0.02
	TE GUI Non-Scrolling Percentage	-6.7	0.56
	TE SUI Non-Scrolling Percentage	6.7	0.56
	Number of Tags Generated	-1.4	0.94
Gender (Male vs. Female)	TE Scrolling Percentage	-14.2	0.28
	TE GUI Non-Scrolling Percentage	9.2	0.43
	TE SUI Non-Scrolling Percentage	-9.2	0.43
	Number of Tags Generated	9.6	0.56
Tagging Event Level (Low vs. High)	TE Scrolling Percentage	-30.7	0.01
	TE GUI Non-Scrolling Percentage	13.6	0.23
	TE SUI Non-Scrolling Percentage	-13.6	0.23
	Number of Tags Generated	-23.9	0.14

Table 4.3.1c. Results of tagging event statistical significance testing

(first set of analytical equations, single-day participants)

Filter	Number of Tags Generated							
Low	20	29	14	35	57	57	22	69
High	31	40	41	36	41	138	73	94

Table 4.3.1d. Number of tags generated under our low/high method of filtering (single-day participants)

4.3.2 Second set of analytical equations

Restated below is the second set of analytical equations of the tagging event logs, the values of which have been depicted in figure 4.3.2 and table 4.3.2a. We find a large number of events from the [Annotating Events] grouping to be events of inserting newly typed tags. The tag insertion percentage (TIP, equation 3.5.14) represents this [Insert_Tag] to [Annotating Events] ratio. The tag duplication and tag bank utilization percentages of equations 3.5.15 and 3.5.16 have been included to reveal a participant's awareness of the tag bank's contents and to distinguish between tags paired with photographs via keyboard and those copied from the tag bank. To better reveal usages we exclude the [Insert_Tag] event from our evaluation of the GUI and SUI non-TIP tagging percentages of equations 3.5.17 and 3.5.18 due to its requirement of keyboard input.

$$\begin{bmatrix}\text{Annotating}\\\text{Events}\end{bmatrix} = \sum \begin{bmatrix}\text{Insert\_Tag}\\\text{Copy\_Tag}\\\text{Remove\_Tag}\\\text{Undo}\end{bmatrix} \qquad \text{Eq. 3.5.13}$$

$$\begin{matrix}\text{Tag Insertion}\\\text{Percentage}\end{matrix}(\text{TIP}) = 100\,\frac{[\text{Insert\_Tag}]}{\begin{bmatrix}\text{Annotating}\\\text{Events}\end{bmatrix}} \qquad \text{Eq. 3.5.14}$$

$$\begin{matrix}\text{Tag Duplication}\\\text{Percentage}\end{matrix} = 100\,\frac{[\text{Insert\_Tag}].\text{Exists\_Tag\_Bank}}{[\text{Insert\_Tag}]} \qquad \text{Eq. 3.5.15}$$

$$\begin{matrix}\text{Tag Bank}\\\text{Utilization Percentage}\end{matrix} = 100\,\frac{[\text{Copy\_Tag}]}{[\text{Copy\_Tag}] + [\text{Insert\_Tag}] - [\text{Insert\_Tag}].\text{Tag\_Bank\_Edit}} \qquad \text{Eq. 3.5.16}$$

$$\begin{matrix}\text{GUI Non-TIP} \\ \text{Tagging Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix}.\text{GUI} - [\text{Insert\_Tag}]}{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix} - [\text{Insert\_Tag}]} \qquad \text{Eq. 3.5.17}$$

$$\begin{matrix}\text{SUI Non-TIP} \\ \text{Tagging Percentage}\end{matrix} = 100\frac{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix}.\text{SUI}}{\begin{bmatrix}\text{Annotating} \\ \text{Events}\end{bmatrix} - [\text{Insert\_Tag}]} \qquad \text{Eq. 3.5.18}$$

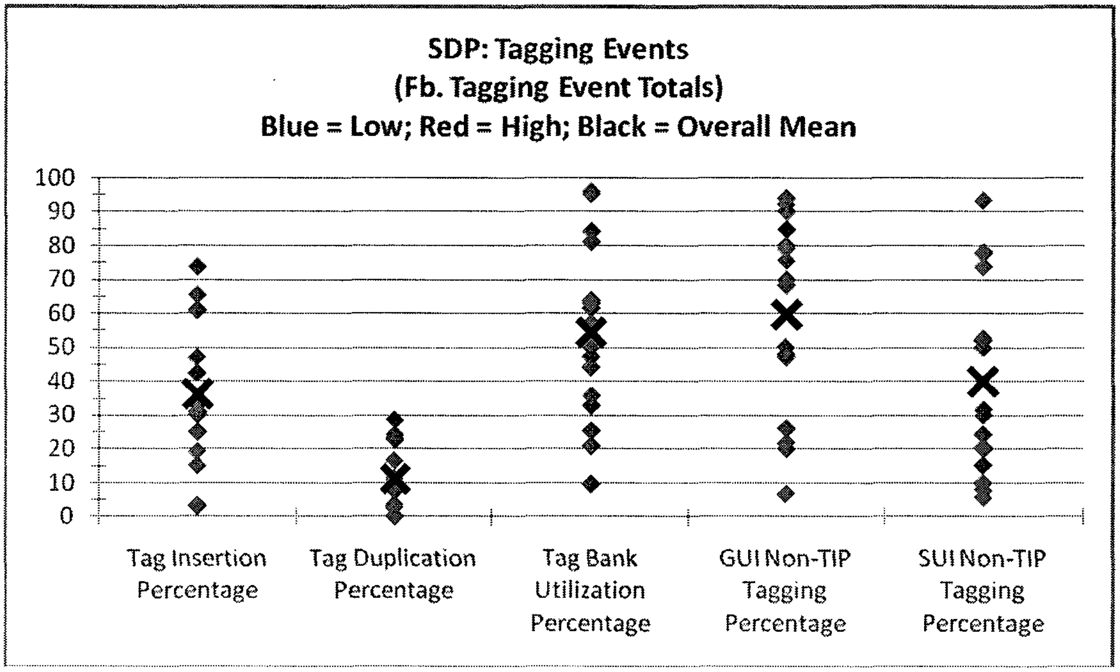


Figure 4.3.2. Characteristics of tagging events filtered by tagging event totals

(second set of analytical equations, single-day participants)

	Filter	Tag Insertion Perc.	Tag Duplication Perc.	Tag Bank Utilization Perc.	GUI Non-TIP Tagging Perc.	SUI Non-TIP Tagging Perc.
Mean	**Low**	46.8	14.7	43.0	62.1	37.6
	High	25.7	7.6	65.6	57.3	42.8
	Overall	36.2	11.1	54.3	59.7	40.2

Table 4.3.2a. Characteristics of tagging events filtered by tagging event totals

(second set of analytical equations, single-day participants)

	Filter	Tag Insertion Perc.	Tag Duplication Perc.	Tag Bank Utilization Perc.	GUI Non-TIP Tagging Perc.	SUI Non-TIP Tagging Perc.
Standard Deviation	Low	23.1	11.1	27.3	21.5	21.0
	High	13.3	6.0	20.4	35.9	35.9
	Overall	21.2	9.3	26.1	28.7	28.5

Table 4.3.2a (continued). Characteristics of tagging events filtered by tagging event totals (second set of analytical equations, single-day participants)

Filtering participants by tagging event totals reveals their tendency to insert newly typed tags at a lesser rate as usage of our application increased (see table 4.3.2b; $p < 0.05$). Along with decreased TIPs, these more ambitious participants also displayed reduced tag duplication percentages (table 4.3.2b; $p = 0.13$). Even further, such participants were found to utilize the tag bank more than those with lowest tagging event totals (table 4.3.2b; $p = 0.08$). These findings, in combination with the significance testing results of table 4.3.1c, where we found participants with high tagging event totals to generate a larger number of tags in comparison to those with low tagging event totals, hint at the idea of proactive tag bank learning. In other words, participants with high tagging event totals scrolled the tag bank more, inserted newly typed tags with the result of duplication less, and in turn, generated more tagging content primarily through tag bank selections.

SUI usage, though, remains somewhat variable. The independence which exists between the variables [Task Completion Rate] and [Tagging Event Level] (see table 4.3.2b; $p > 0.05$) suggests the presence of outliers; since, intuitively, as participants had become more familiar with the SUI of our application, we would expect their task completion rates to have increased – our significance tests, however, do not seem to support this idea.

Independent Variable	Dependent Variables	Mean Difference	Significance
Keyboard Type (QWERTY vs. T9)	Tag Insertion Percentage	16.3	0.13
	Tag Duplication Percentage	3.8	0.43
	Tag Bank Utilization Percentage	-16.7	0.21
	GUI Non-TIP Tagging Percentage	-16.4	0.27
	SUI Non-TIP Tagging Percentage	16.1	0.27
	Task Completion Rate	18.2	0.20
Gender (Male vs. Female)	Tag Insertion Percentage	-11.0	0.32
	Tag Duplication Percentage	-1.6	0.75
	Tag Bank Utilization Percentage	10.8	0.43
	GUI Non-TIP Tagging Percentage	15.8	0.29
	SUI Non-TIP Tagging Percentage	-16.0	0.28
	Task Completion Rate	-1.8	0.90
Tagging Event Level (Low vs. High)	Tag Insertion Percentage	21.0	0.04
	Tag Duplication Percentage	7.1	0.13
	Tag Bank Utilization Percentage	-22.6	0.08
	GUI Non-TIP Tagging Percentage	4.9	0.75
	SUI Non-TIP Tagging Percentage	-5.1	0.73
	Task Completion Rate	-16.5	0.18

Table 4.3.2b. Results of tagging event statistical significance testing (second set of analytical equations, single-day participants)

4.4 Evaluation and discussion of the audio recording logs

4.4.1 Determining outliers

In testing our hypothesis on the voice command usage-task completion rate correlation, we find significance testing across all participants to be an unfair measure due to the following equivalent ratios/non-equivalent parts reasoning: a participant who produced ten correct recognitions, for example, from their total of one hundred SUI utterances has an equivalent task completion rate to another participant who produced one hundred correct recognitions from their total of one thousand SUI utterances. Therefore, our method of low/high filtering based on task completion rate would incorrectly situate our less ambitious SUI participants, those who used voice commands much less, among our more ambitious SUI participants, those with increased usage.

Table 4.4.1a demonstrates this reasoning, showing a low/high filtering of each participant's number of correct recognitions/total number of utterances for control actions and tagging events combined (their overall task completion rate). If we had neglected the above reasoning in the testing of our hypothesis and performed a low/high filtering on task completion rate across all participants, then, as shown in table 4.4.1a, our significance tests would have treated participants with fewest utterances, which just happened to be correctly recognized, alike participants who experienced a much greater number of correct recognitions (e.g. participant 16, alike participant 11, both filtered as having [High] task completion rates). The participants' trying of the SUI modality, though, had very different extents (extreme short-term use/extreme long-term use), so we must not group them in this way.

Filter	Participant	Number of Correct Recognitions/Total Number of Utterances
Low	1	18/57
	2	14/39
	3	13/30
	4	7/16
	5	32/58
	6	36/65
	7	46/83
	8	10/17

Table 4.4.1a. Number of correct recognitions/total number of utterances for control actions and tagging events combined filtered by task completion rate (single-day participants)

Filter	Participant	Number of Correct Recognitions/Total Number of Utterances
High	9	9/15
	10	90/140
	11	82/115
	12	5/7
	13	105/144
	14	11/14
	15	16/17
	16	5/5

Table 4.4.1a (continued). Number of correct recognitions/total number of utterances for control actions and tagging events combined filtered by task completion rate (single-day participants)

Figure 4.4.1, ordered identically to table 4.4.1a, illustrates that many participants simply did not desire to pursue the SUI. Each marking within the figure represents a single utterance issued over the participant's scaled-to-unity timeframe of application use, with black colored markers corresponding to participants with weakest SUI history and red colored markers corresponding to those with strongest SUI history (strength of SUI history based on the median of the group's average scaled-to-unity time between utterance positioning – in table 4.4.1b we list the average scaled-to-unity time between utterance positioning for all participants). Participants with weakest history, those below the median, are shown to exhibit a clustered SUI usage, a result, in our opinion, of their lack of interest with the SUI and not because of poor recognition performance. Those with strongest history, those above the median, exhibit a more distributed SUI usage, implying an interest in the SUI, the extent of which may or may not have been influenced by their task completion rate.

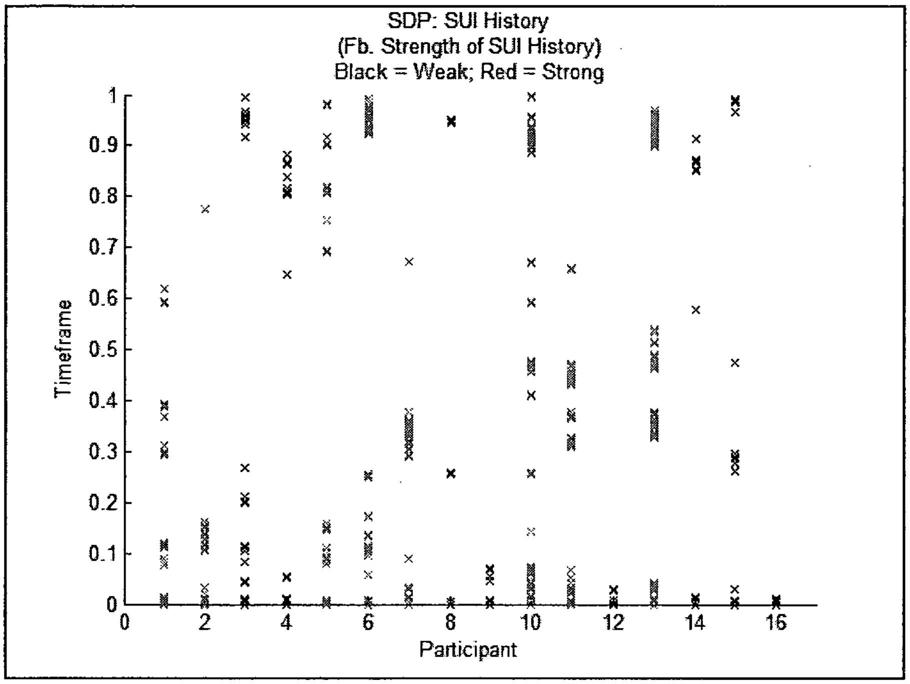


Figure 4.4.1. Timeframe of application use showing utterance positioning (single-day participants)

Participant	1	2	3	4	5	6	7	8
Average Time	0.02	0.03	0.03	0.06	0.02	0.02	0.01	0.06

Participant	9	10	11	12	13	14	15	16
Average Time	0.07	0.01	0.01	0.14	0.01	0.07	0.06	0.25

Table 4.4.1b. Average scaled-to-unity time between utterance positioning (single-day participants)

In the remainder of this section we evaluate upon a reduced participant set, the set of participants with strongest SUI history. We believe that this partitioning will not bias the results of our significance tests to come since, as shown in table 4.4.1c, a good variability in the total number of utterances still exists for the reduced set. In performing this separation we have simply removed those participants from table 4.4.1a who demonstrated an unwillingness to invoke the SUI or pursued the tagging of photographs in settings which deterred their usage of voice commands.

Filter	Participant	Number of Correct Recognitions/Total Number of Utterances
Low	1	18/57
	2	14/39
	5	32/58
	6	36/65
High	7	46/83
	10	90/140
	11	82/115
	13	105/144

Table 4.4.1c. Number of correct recognitions/total number of utterances for control actions and tagging events combined filtered by task completion rate (reduced participant set, single-day participants)

4.4.2 Analytical equations and classification table

The analytical equations and classification table used in evaluating a participant's level of success in SUI usage are restated below. Recalling from chapter 3, section 3.5.3, recognition rate (equation 3.5.21) is a sort of "could have been/would have been" performance measure, while task completion rate (equation 3.5.22) is a performance measure of true participant experiences. Recognition rate, an inflated version of task completion rate, removes seven of nine classifications from the [All Classifications] grouping in an attempt to account for user error.

$$\begin{bmatrix}\text{All}\\\text{Classifications}\end{bmatrix} = \sum \begin{bmatrix}\text{Correct Recognition}\\\text{Out-of-Grammar}\\\text{Substitution}\\\text{Segmentation}\\\text{Deletion}\\\text{Order Rearrangement}\\\text{Disfluency}\\\text{Noisy Environment}\\\text{Unrecognized}\end{bmatrix} \qquad \text{Eq. 3.5.20}$$

$$\text{Recognition Rate} = 100 \frac{[\text{Correct Recognitions}]}{[\text{Correct Recognitions}] + [\text{Unrecognized}]} \qquad \text{Eq. 3.5.21}$$

$$\text{Task Completion Rate} = 100 \frac{[\text{Correct Recognitions}]}{[\text{All Classifications}]} \qquad \text{Eq. 3.5.22}$$

Classification	Description	Incorrect Saying Example (Correct Saying: "remove tag river bed")
Correct Recognition	Properly recognized utterance	N/A
Out-of-Grammar	Grammar does not support utterance	N/A
Substitution	Additional wording within utterance	"remove the tag river bed"
Segmentation	Incomplete utterance	"remove tag river"
Deletion	Necessary words removed from utterance	"remove river bed"
Order Rearrangement	Incorrect ordering within utterance	"river bed remove tag"
Disfluency	Poor articulation within utterance	"remove tag riv bed"
Noisy Environment	Noise overtakes utterance	N/A
Unrecognized	Utterance supported by grammar yet unrecognized	N/A

Table 3.5.3. Classifications of audio recordings, their descriptions and saying examples

4.4.3 Recognition and task completion for control actions

As hypothesized, task completion rate strongly influences a participant's willingness to issue control action voice commands. Differences between the CA GUI and SUI non-scrolling percentages of participants with low and high task completion rates were found to be statistically significant with $p < 0.05$ (table 4.4.3b); in other words, a favoring of the GUI modality for participants with low task completion rates was found to exist.

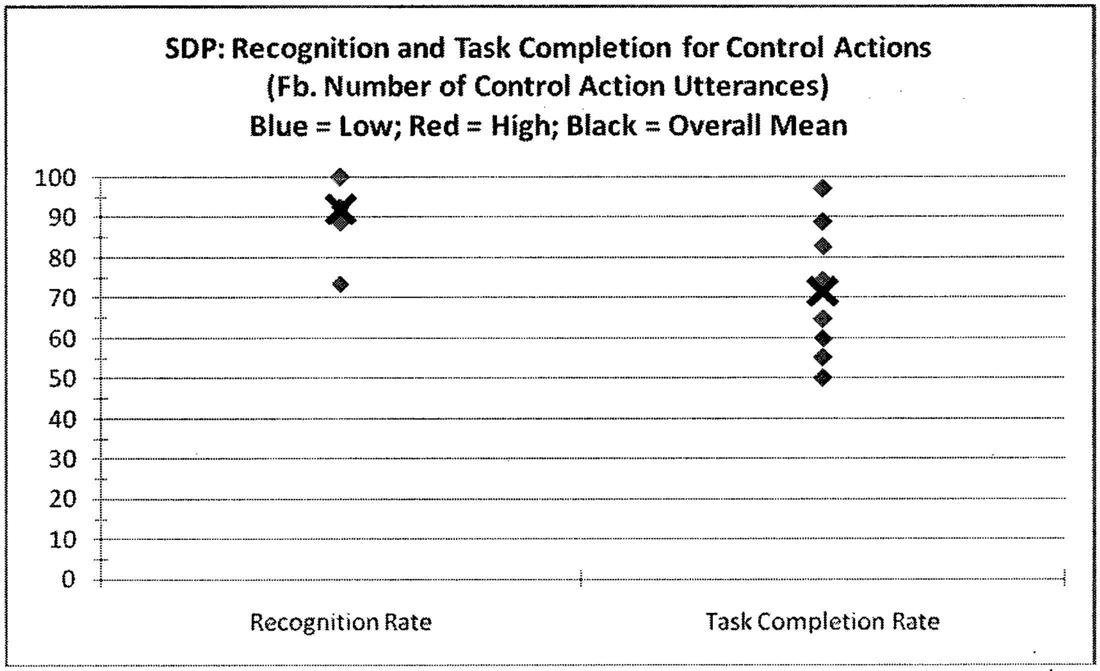


Figure 4.4.3a. Characteristics of recognition and task completion for control actions filtered by number of control action utterances (reduced participant set, single-day participants)

		Recognition Rate	Task Completion Rate
Mean	**Low**	89.5	63.5
	High	94.4	79.7
	Overall	92.0	71.6
Standard Deviation	**Low**	11.4	17.4
	High	6.5	13.8
	Overall	9.0	16.9

Table 4.4.3a. Characteristics of recognition and task completion for control actions filtered by number of control action utterances (reduced participant set, single-day participants)

Independent Variable	Dependent Variables	Mean Difference	Significance
Task Completion Level (Low vs. High)	CA Scrolling Percentage	-3.7	0.77
	CA GUI Non-Scrolling Percentage	20.6	0.05
	CA SUI Non-Scrolling Percentage	-20.6	0.05
	SUI Selections Percentage	-27.2	0.25
	Number of Captured Photographs	-8.3	0.01
	Task Completion Rate	-28.3	0.01

Table 4.4.3b. Results of control actions statistical significance testing

(reduced participant set, single-day participants)

Note: Significance between the variables [Number of Captured Photographs] and [Task Completion Level] to be discussed in section 4.4.4.

From figure 4.4.3b and table 4.4.3c, a filtering on task completion rate for control actions reveal lower task completion rates to be caused first, second and third most from voice commands being issued under the conditions of noisy environment (18.8 %), disfluency (9.6 %) and unrecognized utterances (7.3 %), respectively. Out-of-grammar utterances, the fourth-most cause of lower task completion rates, account for only 3.7 % of participant utterances.

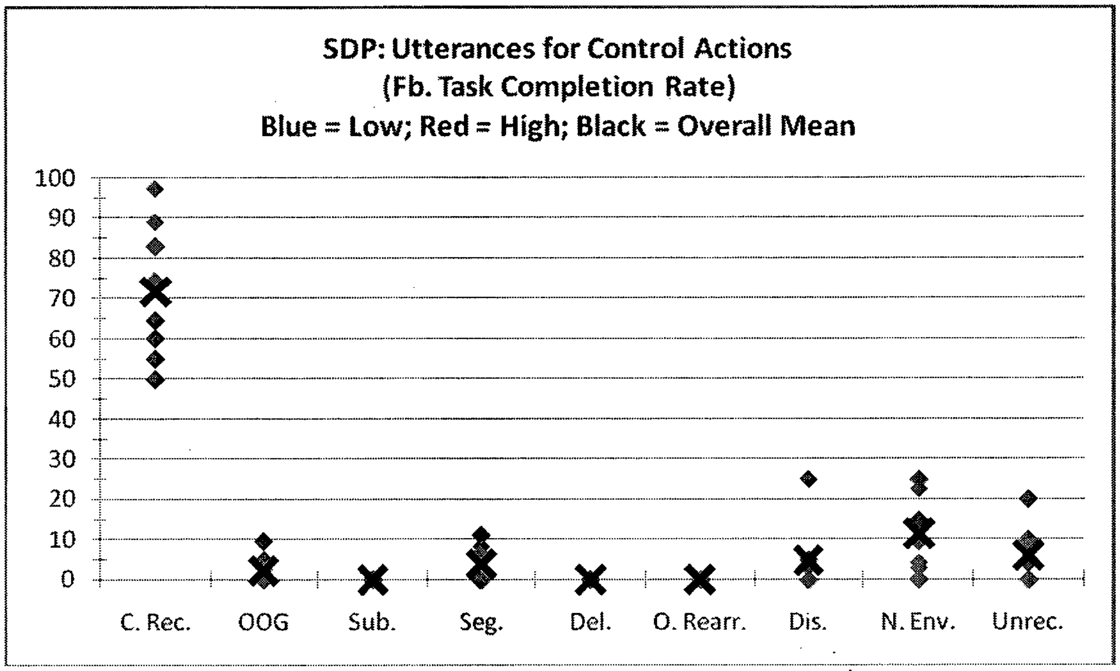


Figure 4.4.3b. Characteristics of utterances for control actions

filtered by task completion rate (reduced participant set, single-day participants)

	Mean			Standard Deviation		
Filter	**Low**	**High**	**Overall**	**Low**	**High**	**Overall**
C. Rec	57.4	85.8	71.6	6.3	9.7	16.9
OOG	3.7	0.7	2.2	4.7	1.4	3.6
Sub.	0.0	0.0	0.0	0.0	0.0	0.0
Seg.	3.3	4.4	3.9	4.1	5.4	4.5
Del.	0.0	0.0	0.0	0.0	0.0	0.0
O. Rearr.	0.0	0.0	0.0	0.0	0.0	0.0
Dis.	9.6	0.0	4.8	10.3	0.0	8.5
N. Env.	18.8	4.2	11.5	6.0	4.1	9.1
Unrec.	7.3	4.9	6.1	8.8	5.7	7.0

Table 4.4.3c. Characteristics of utterances for control actions

filtered by task completion rate (reduced participant set, single-day participants)

4.4.4 Recognition and task completion for tagging events

As for control actions, we also find task completion rate to strongly influence a participant's willingness to issue tagging event voice commands. Differences between the TE GUI and SUI non-scrolling percentages and GUI and SUI non-TIP tagging percentages of participants with low and high task completion rates were found to be statistically significant with $p \ll 0.05$ (table 4.4.4b).

From table 4.4.3b, we found a strong dependence ($p \ll 0.05$) between the variables [Number of Captured Photographs] and [Task Completion Level]. In light of this relationship, and recalling the dependence between [Number of Tags Generated] and [Tagging Event Level] from section 4.3.1, we propose an additional significance test between the variables [Number of Tagging Events] and [Task Completion Level] to determine if task completion could be the driving factor behind our participants' ambition. To clarify:

1. We found participants of lower task completion rates to capture fewer photographs.
2. This reduction in the number of photographs captured, if large, could ultimately diminish the number of tags the participant generates because each tag the participant is said to generate is done so by pairing it with a photograph. A fewer number of photographs to pair tags with may result in fewer tags being paired.
3. With the number of tags generated being dependent upon a participant's tagging event level (which we have rephrased as their ambition), then, perhaps, a participant's task completion rate drives their ambition.

Table 4.4.4b rejects this assumption, and informs us that participants enduring poor SUI performance compensated through GUI interactions (no significance between the variables [Number of Tagging Events] and [Task Completion Level]; $p = 0.76$).

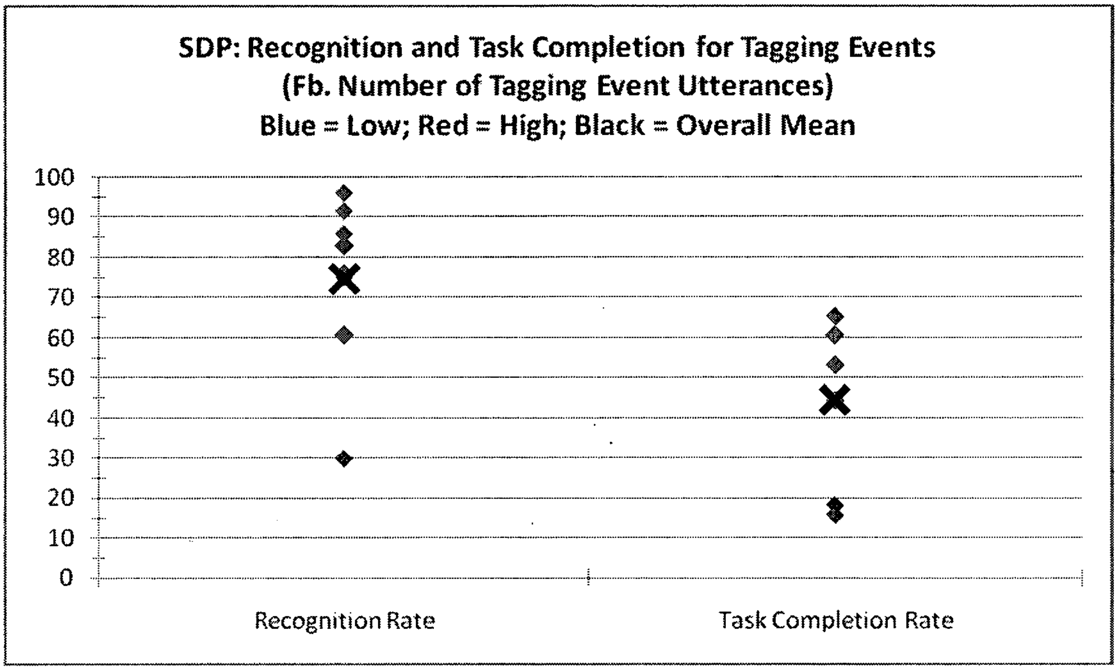


Figure 4.4.4a. Characteristics of recognition and task completion for tagging events filtered by number of tagging event utterances (reduced participant set, single-day participants)

		Recognition Rate	Task Completion Rate
Mean	**Low**	68.4	32.9
	High	80.9	55.9
	Overall	74.6	44.4
Standard Deviation	**Low**	26.0	18.8
	High	16.1	9.3
	Overall	21.1	18.4

Table 4.4.4a. Characteristics of recognition and task completion for tagging events filtered by number of tagging event utterances (reduced participant set, single-day participants)

Independent Variable	Dependent Variables	Mean Difference	Significance
Task Completion Level (Low vs. High)	TE Scrolling Percentage	10.1	0.46
	TE GUI Non-Scrolling Percentage	39.5	0.01
	TE SUI Non-Scrolling Percentage	-39.5	0.01
	Number of Tags Generated	-21.5	0.46
	Tag Insertion Percentage	3.1	0.86
	Tag Duplication Percentage	4.1	0.53
	Tag Bank Utilization Percentage	-6.8	0.75
	GUI Non-TIP Tagging Percentage	46.2	0.01
	SUI Non-TIP Tagging Percentage	-45.7	0.01
	Task Completion Rate	-27.5	0.02
	Number of Tagging Events	-73	0.76

Table 4.4.4b. Results of tagging events statistical significance testing (reduced participant set, single-day participants)

A filtering on task completion rate for tagging events reveal lower task completion rates to be caused first, second and third most from voice commands being issued under the conditions of out-of-grammar (24.2 %; up 20.5 % from control actions), unrecognized utterances (21.0 %; up 13.7 %) and noisy environment (11.8 %), respectively (figure 4.4.4b and table 4.4.4c). Such an increase in the percentage of out-of-grammar utterances places first blame upon a poorly initialized tag bank, while we believe the reduced computational throughputs of our mobile devices to contribute to the increase of unrecognized utterances, especially since the larger grammar of the Tag Bank window, containing up to 225 entries, had been under consideration here.

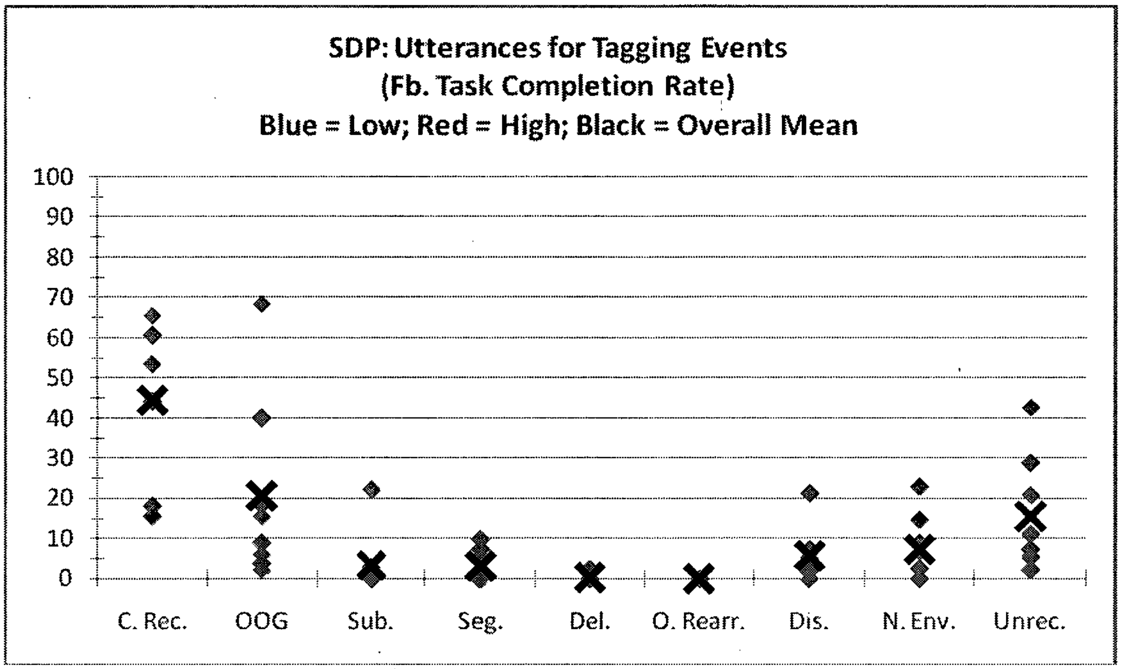


Figure 4.4.4b. Characteristics of utterances for tagging events filtered by task completion rate (reduced participant set, single-day participants)

	Mean			Standard Deviation		
	Low	**High**	**Overall**	**Low**	**High**	**Overall**
C. Rec	30.7	58.2	44.4	15.8	5.9	18.4
OOG	24.2	16.7	20.5	30.2	16.5	22.9
Sub.	0.8	5.8	3.3	1.5	11.0	7.7
Seg.	3.2	3.0	3.1	3.8	4.7	3.9
O. Rearr.	0.0	0.6	0.3	0.0	1.2	0.9
Del.	0.0	0.0	0.0	0.0	0.0	0.0
Dis.	8.5	2.7	5.6	9.1	1.2	6.7
N. Env.	11.8	3.0	7.4	9.7	2.8	8.1
Unrec.	21.0	10.0	15.5	17.8	8.1	14.1

Table 4.4.4c. Characteristics of utterances for tagging events filtered by task completion rate (reduced participant set, single-day participants)

As a comparison, Turner and Kun's [22] analysis of the P54 command-and-control SUI (structurally equivalent to our control action SUI) was found to peak with accuracy of

94.0 % and average at 85.3 %. In our work we've managed an overall (control action/tagging event SUI combination) peak correct recognition rate of only 72.9 % and average of 55.3 %.

One reason for such poor performance comes from an observation made during transcription of the audio recording logs, where we observed some participants pursuing a guess-and-check method to tagging photographs via SUI. Typically, such behavior resulted in an absent or incorrect system response and lead to user frustration (the tone of our participants' in subsequent utterances reflected irritation). This observation, and the fact that 78.0 % of all tag-related utterances were issued from the Tag Bank window, where participants could view the contents of the tag bank, stress the necessity of a visual feedback regarding the grammar's (tag bank's) contents. In our opinion, participants of our experiment actively sought the visual feedback provided by the Tag Bank window.

4.5 Evaluation of the pre-experiment questionnaire

Table 4.5a lists the mean and standard deviation of numerical responses to the pre-experiment questionnaire. Table 4.5b lists the two most common responses, and their percentages, to the Likert scale-based portions of the pre-experiment questionnaire. We postpone discussion on these results to section 4.7.

Question	Mean	Standard Deviation
Age	20.7	1.9
How many pictures do you take on a monthly basis with a digital camera	34.8	32.9
Number of years you've owned a mobile phone	5.4	2.0

Table 4.5a. Numeric responses to the pre-experiment questionnaire (single-day participants)

Question	Mean	Standard Deviation
How many pictures do you take on a monthly basis with a camera phone	7.9	8.1
How many pictures do you view on photo-sharing websites on a monthly basis	162.8	263.2

Table 4.5a (continued). Numeric responses to the pre-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
Education	Sophomore	43.8	Senior	25.0	68.8
I have used digital cameras on multiple occasions	Strongly Agree	87.5	Agree	6.3	93.8
I own a digital camera	Yes	87.5	No	12.5	100.0
I own a mobile phone	Yes	100.0	No	0.0	100.0
My mobile phone has a full/slide-out keyboard	Yes	62.5	No	37.5	100.0
My mobile phone uses T9 text entry	Yes	75.0	No	25.0	100.0
I often type on my mobile phone	Strongly Agree	75.0	Agree	25.0	100.0
I use a stylus to type on my mobile phone	Strongly Disagree	87.5	Neither agree nor disagree	6.3	93.8
I use a full/slide-out keyboard to type on my mobile phone	Strongly Agree	50.0	Strongly Disagree	37.5	87.5
I often send text messages from my mobile phone	Strongly Agree	87.5	Agree	6.3	93.8

Table 4.5b. Likert scale-based responses to the pre-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
I own a camera phone	Yes	93.8	No	6.3	100.0
I'm familiar with desktop/internet photo managers which allow me to tag images	Strongly Agree	56.3	Agree	37.5	93.8
I tag my pictures	Strongly Agree	56.3	Agree	25.0	81.3
Tagging pictures is useful	Strongly Agree	56.3	Agree	31.3	87.5
I currently have an account with a photo-sharing website	Yes	100.0	No	0.0	100.0
Please list websites and year/month you opened the account	Facebook	93.8	Flickr	12.5	106.3 (participants listed multiple websites)
I upload my pictures to the above website(s)	Strongly Agree	68.8	Neither agree nor disagree	18.8	87.5
How often do you visit photo-sharing websites	Daily	50.8	Weekly	50.0	100.0
Please list websites	Facebook	87.5	Flickr	12.5	100.0
I view the tags of the pictures from previous question	Agree	37.5	Strongly Agree	31.3	68.8
I have experience using a PDA, Smartphone, iPhone or similar device	Neither agree nor disagree	37.5	Strongly Agree	25.0	62.5

Table 4.5b (continued). Likert scale-based responses to the pre-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
My PDA, Smartphone, iPhone or similar device allows me to upload pictures to a photo-sharing website	Strongly Disagree	43.8	Neither agree nor disagree	18.8	62.5
I have used voice commands to interact with my PDA, Smartphone, iPhone or similar device	Strongly Disagree	68.8	Neither agree nor disagree	12.5	81.3
Please give examples to the previous question	Search Google	6.3	First response was only response	X	X
I have used speech to interact with computers in the past	Strongly Disagree	56.3	Neither agree nor disagree	18.8	75.0
Please give examples to the previous question	I tried it on my PC	25.0	First response was only response	X	X

Table 4.5b (continued). Likert scale-based responses to the pre-experiment questionnaire (single-day participants)

4.6 Evaluation of the post-experiment questionnaire

Table 4.6 lists the two most common responses, and their percentages, to the Likert scale-based portions of the post-experiment questionnaire. We postpone discussion on these results to section 4.7.

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
I found the capturing of pictures to be simple and intuitive	Strongly Agree	56.3	Agree	31.3	87.5
I found the tagging of pictures to be simple and intuitive	Agree	43.8	Strongly Agree	37.5	81.3
I tagged most of the pictures I took during the experiment	Strongly Agree	87.5	Agree	12.5	100.0
I understand how to tag photographs using voice commands	Strongly Agree	62.5	Agree	31.3	93.8
I found using voice commands to be more convenient than typing when tagging pictures	Neither agree nor disagree	37.5	Disagree	31.3	68.8
I understand the concept of the tag bank	Strongly Agree	68.8	Agree	25.0	93.8
I found the tag bank to be helpful when reusing previously entered tags	Strongly Agree	62.5	Disagree	18.8	81.3
After the initial setup of the application the tag bank contained most of the tags I needed	Disagree	37.5	Strongly Agree	25.0	62.5

Table 4.6. Likert scale-based responses to the post-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
After the initial setup of the application I removed many tags from the tag bank	Strongly Disagree	56.3	Disagree	31.3	0.0
It was easy to use the keyboard	Strongly Agree	62.5	Disagree	18.8	0.0
It was easy to type tags	Strongly Agree	56.3	Agree	18.8	0.0
I used the keyboard to tag most of the pictures	Agree	37.5	Strongly Agree	31.3	0.0
I used the stylus to tag most of the pictures	Strongly Disagree	37.5	Agree	31.3	0.0
The imaging application responded correctly to my tactile commands	Agree	56.3	Strongly Agree	0.0	25.0
The imaging application responded correctly to my voice commands	Agree	37.5	Neither agree nor disagree	0.0	37.5
The "Talk" button was easy to operate	Strongly Agree	56.3	Neither agree nor disagree	0.0	18.8
The imaging application recognized most of my utterances	Strongly Agree	31.3	Neither agree nor disagree	0.0	#VALUE!
Using voice commands was frustrating	Agree	25.0	Neither agree nor disagree	0.0	25.0

Table 4.6 (continued). Likert scale-based responses to the post-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
Using voice commands worked well outdoors	Neither agree nor disagree	37.5	Agree	31.3	68.8
Using voice commands worked well indoors	Strongly Agree	37.5	Agree	37.5	75.0
Using voice commands worked well in noisy environments	Neither agree nor disagree	31.3	Disagree	31.3	62.5
I was satisfied with the quality of the camera	Agree	43.8	Strongly Agree	37.5	81.3
The GUI buttons and fields were large enough to allow easy interaction with the application	Agree	43.8	Strongly Agree	37.5	81.3
The tutorial(s) detailed the imaging application in a clear manner	Strongly Agree	62.5	Agree	25.0	87.5
I had to reference the tutorial(s) many times while using the imaging application	Strongly Disagree	62.5	Disagree	37.5	100.0
I enjoyed participating in this experiment	Strongly Agree	62.5	Agree	37.5	100.0
I liked today's weather	Strongly Agree	43.8	Agree	25.0	68.8

Table 4.6 (continued). Likert scale-based responses to the post-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
Today was a difficult day at school and/or work	Neither agree nor disagree	37.5	Disagree	0.0	31.3
Tagging pictures is useful	Agree	56.3	Strongly Agree	0.0	31.3
Tagging pictures immediately after they are taken is useful	Agree	56.3	Strongly Agree	0.0	25.0
I would install a free camera phone application that allowed me to tag pictures using voice commands and upload them to a photo-sharing website	Agree	50.0	Strongly Agree	0.0	25.0
I would install a free camera phone application that allowed me to tag pictures using a GUI (but not voice commands) and upload them to a photo-sharing website	Agree	50.0	Neither agree nor disagree	25.0	0.0
I would pay for a camera phone application that allowed me to tag pictures using voice commands and upload them to a photo-sharing website	Disagree	50.0	Neither agree nor disagree	25.0	0.0

Table 4.6 (continued). Likert scale-based responses to the post-experiment questionnaire (single-day participants)

Statement	First Response	Percentage	Second Response	Percentage	Percentage Sum
I would pay for a camera phone application that allowed me to tag pictures using a GUI (but not voice commands) and upload them to a photo-sharing website	Disagree	56.3	Neither agree nor disagree	25.0	81.3

Table 4.6 (continued). Likert scale-based responses to the post-experiment questionnaire (single-day participants)

4.7 Addressing poor task completion

In this section of the work we would like to address our main concern: poor task completion. Individually, each prior evaluation has hinted at factors which may support the inconsistency; therefore, we will build upon this foundation by extending, reformatting and reviewing some of our most promising findings. Participant responses from the pre- and post-experiment questionnaires will be used to further strengthen our quantitative results.

Participants of this experiment seemed to have not chosen the SUI as their modality preference, a consequence, we feel, of the participants' lack in familiarity with the SUI's design and the initialization procedures of the tag bank. Reviewing the results of our evaluations we first highlight lack in familiarity with design. We observed many participants as demonstrating a proactive learning of the tag bank's contents; that is, our more ambitious participants were found to scroll the tag bank more, insert newly typed tags with the result of duplication less and generate more tagging content, primarily through GUI tag bank selections, as their usage evolved. Learning, though, takes time. Dedicated, but inexperienced, SUI users issuing, at first, a significant percentage of user-in-error utterances, delay their task completion from its true potential at end and allow less ambitious SUI users to produce task completion rates comparable to or in excess of their own (recall from section 4.4.1, our strength of SUI history measure based on a participant's average scaled-to-unity time between utterance positioning). With this in mind we turn to the pre-experiment questionnaire, where we find a strong handling of digital cameras and mobile phones but a lesser handling of Pocket PC and Smartphone devices. Even though each participant was first trained in issuing voice commands through the on-device tutorial, we see this lack in practice with newer technologies as contributing negatively towards voice command usage and task completion on the mobile device. A prior-to-experiment learning of the tag bank's contents and a strong understanding of how to use voice commands would make otherwise unknown tags available to the user for SUI selection and potentially increase their dedication to the modality along with their task completion rate.

Aside from familiarity with design, we believe the second most reputable cause of poor task completion and reduced SUI usage to lie in our methods of tag bank initialization. Figure 4.7a shows participants with highest SUI non-TIP tagging percentages as still demonstrating mixed TIPs, and thus implying a poorly initialized tag bank. The overall means of figure 4.7b confirm this finding, showing that well over half of the tags utilized by all participants were not initially provided, they were inserted (percent generated from insertion mean: 65.3 %, percent generated from interest mean: 32.9 %, percent generated from account mean: 1.8 %). In our first method of tag bank initialization, tags related to particular interests of our participants, generated by users of Flickr with similar interests, were extracted from Flickr's database and inserted into the tag bank for later use. Since the interests specified by our participants were not recorded we are unable draw conclusions upon how well each related to their captured photographs; yet, intuitively, a poor relationship between our participants' interests and their captured photographs necessitate tag insertions. The fact that out-of-grammar utterances rose 20.5% from control actions to tagging events show our user-interest method of tag bank initialization as unfulfilling to the participants' needs.

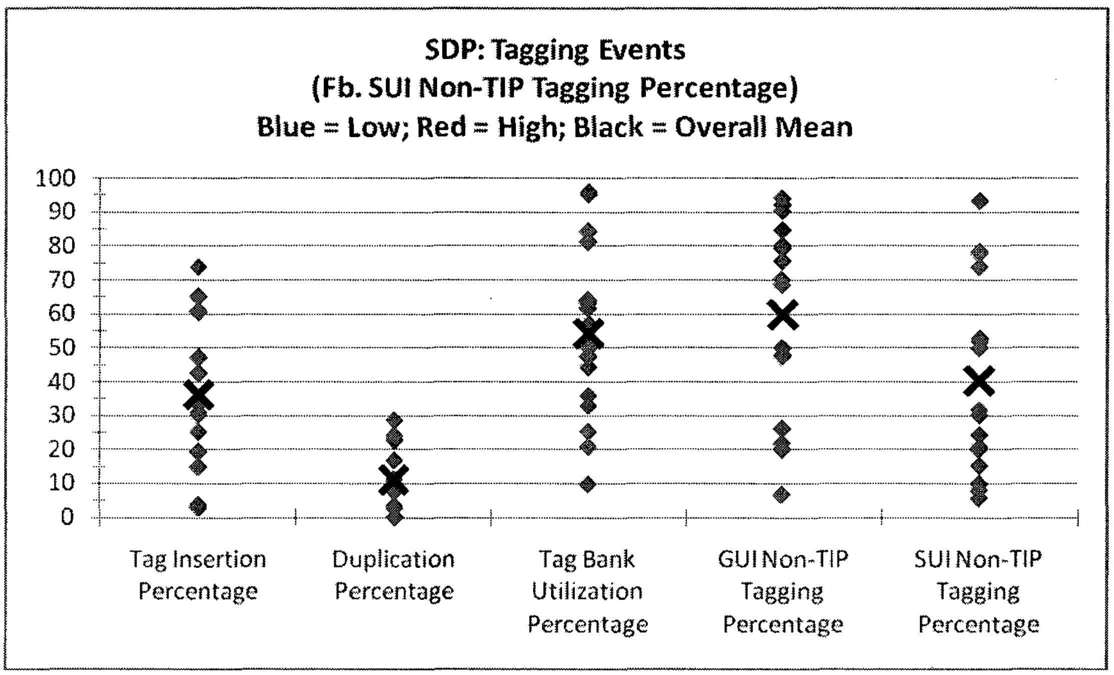


Figure 4.7a. Characteristics of tagging events

filtered by SUI non-TIP tagging percentage (single-day participants)

A special case does exist; specifically, the only participant with a preexisting Flickr account (figure 4.7b, percent generated from account: 28.6 %). This participant was among lowest in TIP and among highest in SUI usage and percentage of correct recognitions for tagging events, a byproduct of, perhaps, our second method of tag bank initialization, where the user's most frequently used tags are extracted from his or her Flickr account and inserted into the tag bank for later use. Extracting tags from this participant's account may have simulated the prior-to-experiment learning of the tag bank's contents we assume has the potential to increase SUI dedication and task completion above participants of a tag bank initialized solely on user interests.

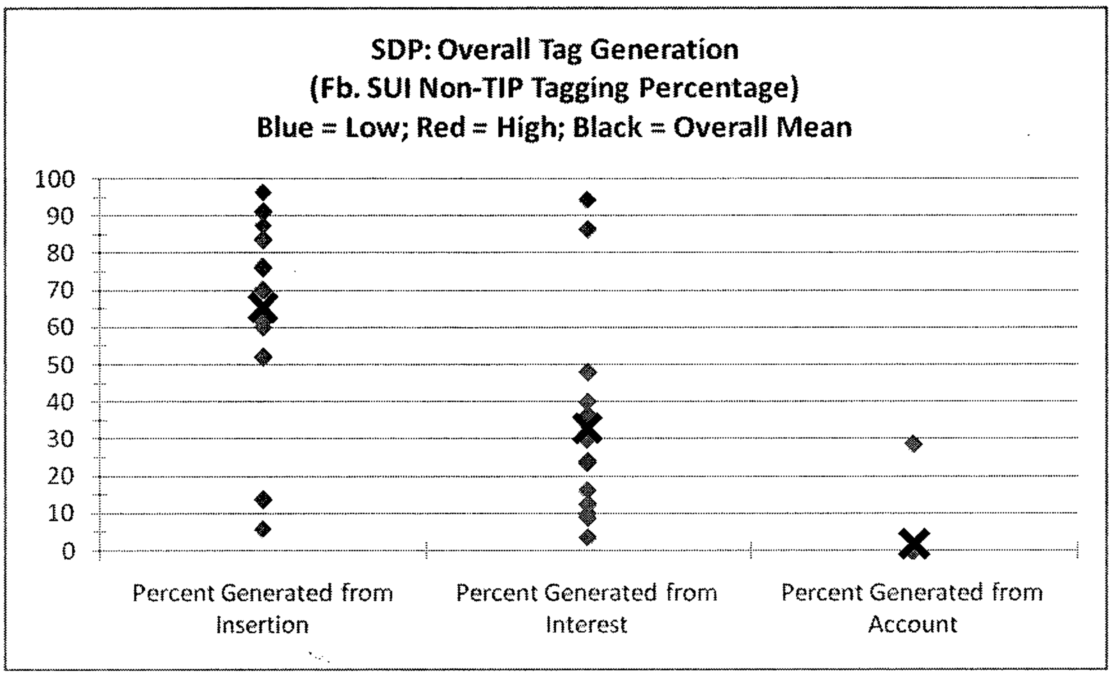


Figure 4.7b. Characteristics of tag generation

filtered by SUI non-TIP tagging percentage (single-day participants)

The sheer size of a freshly initialized tag bank grammar, ranging up to 225 entries, may have exceeded the computational power of the mobile device's processor. From the post-experiment questionnaire, we find that participant's did not remove tags from the tag bank following initialization. As new tags were inserted the size of the grammar expanded. Recalling a prior analysis, where we found unrecognized utterances to increase 13.7 % from control actions to tagging events, gives indication of processor stress.

The overpowering tag insertion percentages we've come across may be reflected in the pre-experiment questionnaire, where participants claim to be quite accustomed to performing QWERTY/T9 text entry in text message compositions. Perhaps our participants simply preferred this type of interaction over GUI and SUI selections. Figure 4.7c below is a reformation of a prior evaluation upon the tagging event logs and filters participants by TIP. This filtering type is synonymous to classifying participants based on their level of keyboard use since each newly inserted tag first requires keyboard input. In doing so we are able to distinguish each participant's keyboard use from the set [Low, Moderate, High] representing the axis partitions [[0:29], [30:59], [60:100]].

Here, we introduce a new variable, the tag existence percentage (TEP, equation 4.7), representing the ratio of [Insert_Tag] events with tags which had existed in the tag bank listing to the total number of events in which a tag from the tag bank had been paired with a photograph.

$$\underset{\text{Percentage}}{\text{Tag Existence}}\ (\text{TEP}) = 100\,\frac{[\text{Insert\_Tag}].\text{Exists\_Tag\_Bank}}{[\text{Copy\_Tag}] + [\text{Insert\_Tag}].\text{Exists\_Tag\_Bank}} \qquad \text{Eq. 4.7}$$

Contrary to our assumption, which situate our participants' preference on inserting newly typed tags, figure 4.7c and table 4.7 show the majority of the participant group as demonstrating low TEPs, implying the group's true preference to be the selection of tags from the tag bank when aware of the tag's existence within it. Again, since the interests specified by our participants were not recorded we are unable draw direct conclusions upon how well each related to their captured photographs. Yet, our high TIP-lower TEP discovery strongly suggests a poor relationship between the interests specified by our participants during tag bank initialization and their captured photographs.

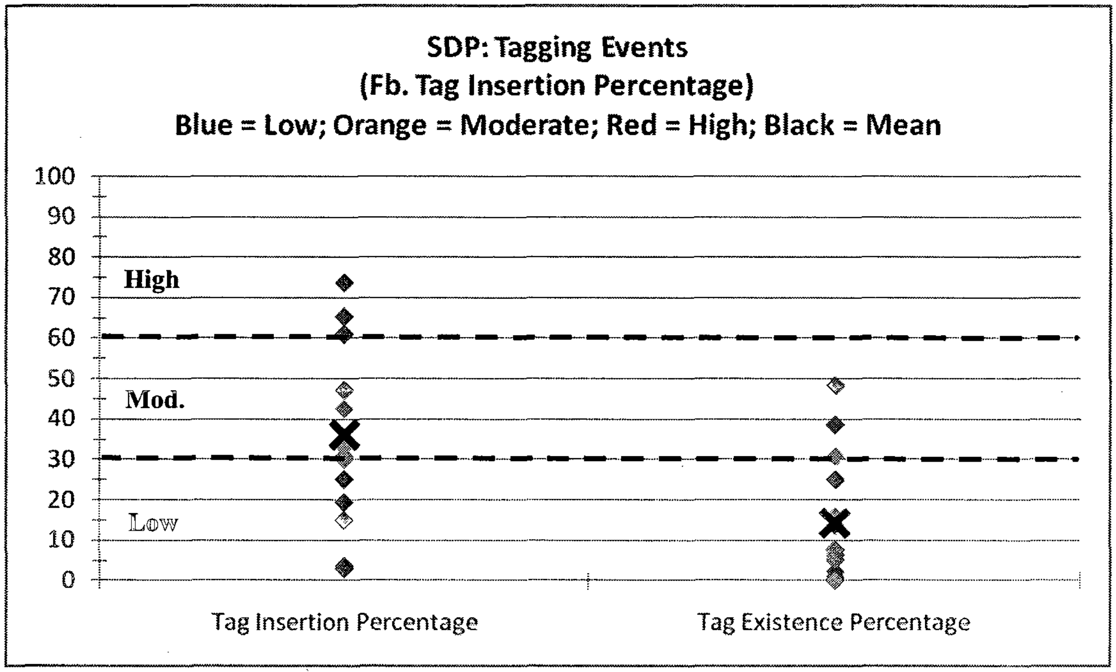


Figure 4.7c. Characteristics of tagging events

filtered by tag insertion percentage (single-day participants)

Usage of Keyboard (TIP)	Existing Tags Usage of Keyboard (TEP)	Number of Participants
Low	Low	5
Moderate	Low	6
Moderate	Moderate	1
High	Moderate	2
High	Low	2

Table 4.7. Classifications of tag generation (single-day participants)

CHAPTER 5

CONCLUSION

5.1 Our work

The problem we address in this work is that, when tagging photographs, the annotation processes of today's mobile tagging applications focus the user's attention inward on the device and off the environment. Our primary goal, to explore methods of tagging photographs which allow users to focus on their environment, was achieved through the creation of our mobile tagging application and data collection server. To validate our hypotheses, we designed a between-subjects experiment in which participants were asked to extend the activities found in the use of a typical digital camera by tagging the photographs they captured. Our qualitative analysis of the study consisted of pre- and post-experiment Likert scale-based questionnaires, and our quantitative analysis evaluated each participant based on the information found within their logging files.

5.2 Our findings

5.2.1 The notion of ambition and SUI history

The notion of ambition, first arising in section 4.2, separates our participants based on the number of control actions and tagging events they generated while involved in our study. Participants with highest totals were said to have an increased ambition towards our application, in the idea that, while capturing a statistically equivalent number of photographs in comparison to participants of lesser ambition, the extent to which participants of higher ambition reviewed previously captured photographs and the information within the text listings of our application proved to be significant. The notion of ambition takes on an even greater meaning in section 4.3, where we found our more ambitious participants to display a proactive learning of the tag bank's contents; that is, they were found to scroll the tag bank more, insert newly typed tags with the result of duplication less and generate more tagging content, primarily through tag bank selections, as their usage evolved.

In section 4.4, we halved our participant set based on the participants' strength of SUI history (a measure of time between their utterance positioning). While all participants of the reduced set, those with strongest SUI history, showed dedication to the SUI, 5 of 8 of these participants were also found to demonstrate the proactive learning behavior. These participants most strongly support our first hypothesis: a convenient way of tagging digital photographs is by using voice commands, for they most thoroughly examined our application and overcame their lack in familiarity with its design. The 3 of 8 participants from the reduced set who did not demonstrate the proactive learning behavior, we feel, would have perceived the SUI to provide a lesser, but still positive, level of convenience since their tag duplication rates and number of tags generated were found to be inferior to other participants of the set, areas which a highly-convenient SUI should avoid.

The task completion rates for the reduced participant set, though, would seem to suggest otherwise. There, we state a peak correct recognition rate of 72.9 % and average of only 55.3 %. How could an interface which fails nearly as much as it succeeds possibly be perceived as convenient? In Section 4.7 we reason that poor task completion may have, at first, originated from our participants' inexperience with SUI technologies; that is, the bias in task completion rate introduced by our participants at experiment start may have delayed their task completion rate from its true potential at experiment end. Although additional work is needed to verify increased task completion at experiment end, our intuition seems to agree with the findings of Cox et al. [23], who suggest that users are prepared to sacrifice task completion and accuracy in favor of a less demanding, which we reasonably rephrase here as more convenient, interface.

> H1. A convenient way of tagging digital photographs is by using voice commands, thereby focusing the user's attention outward on the environment and possibly resulting in enriched observations (in comparison to the observations made when tagging photographs at a later time)
>
> **Supported, and open to future work**
>
> **Observation enrichment not evaluated**

5.2.2 An improperly initialized tag bank

With the number of out-of-grammar utterances for tagging events rising 20.5 % from control actions, we place our first blame for reduced task completion upon a poorly initialized tag bank. Such an increase casts the tag bank as unfulfilling to the participant's needs, and in section 4.7, we confirm this assumption, finding that for even our most active SUI users, mixed tag insertion rates were not uncommon, and that well over half of the tags utilized by all participants were not initially provided, they were inserted. In our first method of tag bank initialization, tags related to particular interests of our participants, generated by users of Flickr with similar interests, were extracted from Flickr's database and inserted into the tag bank for later use. Since the interests specified by our participants were not recorded we are unable draw conclusions upon how well each related to their captured photographs; yet, intuitively, a poor relationship between our participants' interests and their captured photographs necessitate new tag insertions.

A special case does exist; specifically, the only participant with a preexisting Flickr account. This participant was among lowest in new tag insertion and among highest in SUI usage and percentage of correct recognitions for tagging events, a byproduct of, perhaps, our second method of tag bank initialization, where the user's most frequently used tags are extracted from his or her Flickr account and inserted into the tag bank for later use. Extracting tags from this participant's account may have simulated a prior-to-experiment learning of the tag bank's contents.

One observation lightens the harshness we place on our user-interest method of initialization. During transcription of the audio recordings logs, we observed some participants pursuing a guess-and-check method to tagging photographs via SUI. Typically, such behavior resulted in an absent or incorrect system response and lead to user frustration (the tone of our participants' in subsequent utterances reflected irritation). This observation, and the fact that 78.0 % of all tag-related utterances were issued from the Tag Bank window, where participants could view the contents of the tag bank, stress the necessity of a visual feedback regarding the grammar's (tag bank's) contents. In our opinion, and viewed as a flaw in application design, participants of our experiment actively sought the visual feedback provided by the Tag Bank window. The low percentage of tag duplications, averaging at 11.1 % across all participants, though, still places the majority of blame for reduced task completion upon poor initialization.

> H2. In order to allow users to tag photographs using speech we should populate the grammar with user-dependent information obtained from two online resources:
>
> a. *Tags generated by the user and attached to photographs on one or more of the user's online photo-sharing accounts*
>
> b. *Tags related to particular interests of the user generated by other users with similar interests*
>
> **H2a: Inconclusive, open to future work**
>
> **H2b: Not supported, open to future work**

5.2.3 Discussing keyboard usage

In section 4.7 we had found that each of our participants were quite accustomed to performing QWERTY/T9 text entry in text message compositions, disallowing the drawing of conclusions upon our third hypothesis: self-identified highly-experienced keyboard users will use voice commands less than those with lower keyboard experience. Later in that section of the work, we pondered: with such high levels of experience, perhaps the overpowering tag insertion percentages (TIPs) we've come across are a result of participants simply preferring this type of interaction over GUI and SUI selections. To test, we introduced a new variable, the tag existence percentage (TEP), representing the ratio of tag insertion events with tags which had existed in the tag bank listing to the total number of events in which a tag from the tag bank had been paired with a photograph. Contrary to our assumption, which situates our participants' preference on inserting newly typed tags, the majority of the participant group was found as demonstrating low TEPs, implying the group's true preference to be the selection of tags from the tag bank when aware of the tag's existence within it. Again, since the interests specified by our participants were not recorded we are unable draw direct conclusions upon how well each related to their captured photographs. Yet, our high TIP-lower TEP discovery strengthens our statement regarding the interests specified by our participants during tag bank initialization and their captured photographs (poorly related).

H3. Self-identified highly-experienced keyboard users will use voice commands less than those with lower keyboard experience, especially if they have only a short period of time (e.g. one day) to use our tagging application

Not Evaluated, all participants highly-experienced

H4. Participants suited with our T9-equipped device will generate a greater number of tags over those suited with the QWERTY-equipped device since the text entry process on our T9 device avoids sliding out the keyboard

Not supported, shown to be not statistically significant but not considered further

5.2.4 Task completion rate

As hypothesized, task completion rate was found to strongly influence a participant's willingness to issue both control action and tagging event voice commands. In section 4.4, when reducing our participant set to exclude those with weakest SUI history, we found the percentage of unrecognized tagging event utterances to increase 13.7 % from that of control actions. As equally an important finding as the out-of-grammar influence upon task completion, this increase in unrecognized utterances revealed a second cause for a reduced completion rates – computational restriction. The sheer size of a freshly initialized tag bank grammar, ranging up to 225 entries, may have exceeded the computational power of the mobile device's processor. As new tags were inserted the size of the grammar expanded, and recalling from section 4.7, we found that participant's did not remove tags from the tag bank following initialization. These new tag insertions, then, were being made upon an already overly-congested tag bank.

Interestingly, though, poor task completion was not found to influence our participant's ambition. In section 4.4, we found participants of lower task completion rates to have captured fewer photographs. This reduction in the number of photographs captured, if large, was thought to diminish the number of tags a participant would generate because each tag a participant is said to generate is done so by pairing it with a photograph – a fewer number of photographs to pair tags with may result in fewer tags being paired. With the number of tags generated found as being dependent upon a participant's ambition, then, perhaps, a participant's task completion rate was the driving factor of their ambition. However, this was found to not be the case, and informs us that participants enduring poor SUI performance compensated with GUI interaction.

H5. A correlation between voice command usage and task completion rate exists; that is, users who unsuccessfully issue voice commands relatively early in the process of learning how to use our tagging application will not keep at it – they will give up on voice commands and pursue other interaction modalities
Supported

CHAPTER 6

SUGGESTIONS FOR FUTURE WORK

6.1 Gathering of the corpus

The gathering of the corpus, considered as one of this work's greatest achievements, not only allowed for the evaluation of our application, but serves as an initial dataset for future evaluations.

6.2 Unrecognized utterances

In our evaluation of the audio recording logs, we found the second most substantial cause of lower task completion rates for tagging events to be unrecognized utterances. We believe the reduced computational throughputs of our mobile devices to have contributed to the number of such utterances, especially since the larger grammar of the Tag Bank window had been under consideration here. To ensure this as the cause, a comparison with a PC-based recognition engine's output, after identical initialization, could be perused.

6.3 Close-talk microphones

The severity of noise within the user's environment, our first most cause of lower task completion rates for control actions and third most cause in tagging events, could be reduced through the use of close-talk microphones/Bluetooth headsets and possibly improve recognition performance.

6.4 Tracking of user interests

In our method of tag bank initialization, tags related to particular interests of the user, generated by other users with similar interests, were extracted from Flickr's database and inserted into the tag bank for later use. Since the interests specified by our participants were not recorded we are unable draw conclusions upon how well each related to their captured photographs. The tracking of user interests may refine our poorly initialized tag bank finding to a poorly chosen set of user interests.

6.5 Enforce an editing of the tag bank's initial contents

Entries of the tag bank compose the grammar and largely define the validity of a user's utterance; in other words, the success of a tagging voice command depends upon the existence of the intended tag in the grammar. In an effort to increase recognition performance, we could enforce an editing of the tag bank's initial contents. A smaller grammar imposes less stress upon the recognition engine and reduces the number of comparisons it must make for user utterances.

6.6 Grammar weighting

A great number of tags were stored in the tag bank, each being assigned an equal probability of selection. Clearly, after the user's selection of even their first tag, all entries of the grammar no longer possess equality in probability. To improve recognition performance, a dynamic weighting scheme could be implemented.

6.7 Visual feedback accommodation

In an effort to avoid tag insertions, some participants seemed to pursue a guess-and-check method of tagging photographs via SUI which may have resulted in their frustration (during transcription of the audio recording logs we observed the tone of our participants' to reflect irritation as a result of an absent or incorrect system response). This observation, and the fact that 78.0 % of all utterances were issued from the Tag Bank window, where participants could view the contents of the tag bank, stress the necessity of a visual feedback regarding the tag bank's contents. Accommodations to this requirement could be made to the structure of our application, possibly reducing the counts of out-of-grammar occurrences and the notion of a faulty interface.

6.8 A special case

The one participant who had an existing Flickr account showed higher ambition towards our application, was among lowest in tag insertions and among highest in SUI usage and percentage of correct recognitions. Extracting tags from a participant's Flickr account may simulate the prior-to-experiment learning of the tag bank's contents we assume has the potential to increase SUI dedication and task completion above participants of a tag bank initialized solely on user interests. More work is needed to confirm.

6.9 Proactive tag bank learning

We observed many participants as demonstrating a proactive learning of the tag bank's contents; that is, our more ambitious participants were found to scroll the tag bank more, insert newly typed tags with the result of duplication less and generate more tagging content, primarily through GUI tag bank selections, as their usage evolved. Learning, though, takes time. We reason that poor task completion may have, at first, originated from our participants' inexperience with SUI technologies; that is, the bias in task completion rate introduced by our participants at experiment start may have delayed their task completion rate from its true potential at experiment end. Additional work is needed to verify increased task completion at experiment end, and our behavioral theory on proactive learning.

LIST OF REFERENCES

[1] Yahoo! Inc. Research Berkeley, "ZoneTag Photos," http://zonetag.research.yahoo.com, 2006.

[2] Critical Path Inc., "ShoZu," http://www.shozu.com, 2010.

[3] Yahoo! Inc., "Flickr," http://www.flickr.com, 2010.

[4] Albert Pelhe, Andrew L. Kun, and W. Thomas Miller III, "Project54 system software architecture," *Winter International Symposium on Information and Communication Technologies*, 2004, pp. 1-6.

[5] Yahoo! Inc., "Flickr API," http://www.flickr.com/services/api, 2010.

[6] W. Thomas Miller III, *Overview of Project54 Inter-Application Messaging: Local and Distributed Messaging Architectures*, Technical report, ECE.P54.2004.11, University of New Hampshire, ECE, Project54, 2004.

[7] Rensis Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, 1932, pp. 1-55.

[8] Mark Weiser, "The computer for the 21st century," *Human-Computer Interaction: Toward the Year 2000,* Morgan Kaufmann Publishers Inc., 1995.

[9] Mark Weiser and John Seely Brown, "The coming age of calm technology," *Beyond Calculation: The Next Fifty Years,* Copernicus, 1997.

[10] Tim Kindberg, Mirjana Spasojevic, Rowanne Fleck and Abigail Sellen, "The ubiquitous camera: an in-depth study of camera phone use," *IEEE Pervasive Computing,* vol. 4, 2005, pp. 42-50.

[11] Nancy A. Van House, "Flickr and public image sharing: distant closeness and photo exhibition," *CHI '07 Extended Abstracts on Human Factors in Computing Systems,* 2007.

[12] Lorant Farkas, Jan Blom and Severi Uusitalo, "Social uses of digital imagery with a focus on smart phones applicability," *CHI '06*, 2006.

[13] Morgan Ames and Mor Naaman, "Why we tag: motivations for annotation in mobile and online media," *CHI '07,* 2007, pp. 971-980.

[14] Cameron Marlow, Mor Naaman, Danah Boyd and Marc Davis, "HT06, tagging paper, taxonomy, Flickr, academic article, to read," *The 17th Conference on Hypertext and Hypermedia,* 2006, pp. 31-40.

[15] Nuno Tomas, Tiago Guerreiro and Daniel Goncalves, "StoryTags: once upon a time, there was a photo," *The 27th International Conference Extended Abstracts on Human Factors in Computing Systems,* 2009, pp. 3967-3972.

[16] Mercan Topkara, Bernice Rogowitz, Steve Wood and Jeff Boston, "Collaborative editing of micro-tags," *The 27th International Conference Extended Abstractions on Human Factors in Computing Systems,* 2009, pp. 4297-4302.

[17] Catherine C. Marshall and A.J. Bernheim Brush, "Exploring the relationship between personal and public annotations," *The 4th ACM/IEEE-CS Joint Conference on Digital Libraries,* 2004, pp. 349-357.

[18] Daniel Vogel and Patrick Baudisch, "Shift: a technique for operating pen-based interfaces using touch," *CHI '07,* 2007, pp. 657-666.

[19] Amy K. Karlson and Benjamin B. Bederson, "One-handed touchscreen input for legacy applications," *CHI '08,* 2008, pp. 1399-1408.

[20] Koji Tatani and Khai N. Truong, "An evaluation of stylus-based text entry methods on handheld devices," *The 9th International Conference on Human Computer Interaction with Mobile Devices and Services,* 2007, pp. 487-494.

[21] Christina L. James and Kelly M. Reischel, "Text input for mobile devices: comparing model prediction to actual performance," *CHI '01,* 2001, pp. 365-371.

[22] Laslo Turner and Andrew L. Kun, "Evaluating the Project54 speech user interface," *Adjunct Proceedings of the Third International Conference on Pervasive Computing,* vol. 191, A. Ferscha, R. Mayrhofer, T. Strang, C. Linnhoff-Popien, A. Dey, A. Butz, A. Schmidt (Eds.), 2005.

[23] Anna L. Cox, Paul A. Cairns, Alison Walton and Sasha Lee, "Tlk or txt? Using voice as input for SMS composition," *Personal and Ubiquitous Computing,* vol. 12, 2008, pp. 567-588.

[24] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith and Pete Steggles, "Towards a better understanding of context and context-awareness,"

The 1st International Symposium on Handheld and Ubiquitous Computing, 1999, pp. 304-307.

[25] Mike Hazas, James Scott and John Krumm, "Location-aware computing comes of age," *Computer,* vol. 37, 2004, pp. 95-97.

[26] Eric W. Pfeiffer, "WhereWare," MIT Technology Review, 2003.

[27] Borkur Sigurbjornsson and Roelof van Zwol, "Flickr tag recommendation based on collective knowledge," *The 17th International World Wide Web Conference,* 2008, pp. 327-336.

[28] Yvonne Rogers, "Moving on from Weiser's vision of calm computing: engaging UbiComp experiences," *Ubicomp 2006*, 2006, pp. 404-421.

[29] Thomas Erickson, "Some problems with the notion of context-aware computing," *Communications of the ACM,* vol. 45, 2002, pp. 102-104.

[30] Michael Farrar, "Using voice to tag digital photographs on the spot: installation," http://www.eceblogger.com/using-voice-to-tag-digital-photographs-on-the-spot-installation (Last accessed 2009).

[31] Tim Paek, Sudeep Gandhe, David Maxwell Chickering and Yun Cheng Ju, "Handling out-of-grammar commands in mobile speech interaction using backoff filler models," *Workshop on Grammar-Based Approaches to Spoken Language Processing,* 2007, pp. 33-40.

APPENDIX A

CONSENT AND RELEASE FORMS

CONSENT FORM FOR PARTICIPATION IN A RESEARCH STUDY
Local/Borrowed Version

TITLE OF RESEARCH STUDY

Using Voice to Tag Digital Photographs on the Spot

Conducted by Michael A. Farrar
Graduate student of the Electrical and Computer Engineering Department at the University of New Hampshire

WHAT IS THE PURPOSE OF THIS STUDY?

The purpose of this research is to determine if voice is an appropriate form of interaction for textual labeling.

WHAT DOES YOUR PARTICIPATION IN THIS STUDY INVOLVE?

You will extend the activities found in the use of a typical digital camera by tagging the photographs you capture on a borrowed cellular phone-like Pocket PC device. Project54 will lend you this device, to which Project54's imaging application will be preinstalled, allowing you to tag photographs on-the-spot via voice commands or through traditional methods of keyboard text entry. As an additional feature, the imaging application will allow for you to upload your photos and tagging content to Flickr, a popular photo-sharing website; however, this is not a requirement of this study. An on-device tutorial will guide you through the imaging application's use. This tutorial is accessible at any time. You will also be provided with documented instructions. Your involvement with this study will last for one week.

WHAT ARE THE POSSIBLE RISKS OF PARTICIPATING IN THIS STUDY?

You are under no risk by participating in this study.

WHAT ARE THE POSSIBLE BENEFITS OF PARTICIPATING IN THIS STUDY?

You will not receive any direct benefits from this study. However, and at a community level, this study will propose an alternative method to the tagging of photographs, removing consumers from the upload-then-tag requirement of the non-mobile desktop-computer environment. In general, this study proposes an alternative to the task of textual labeling.

IF YOU CHOOSE TO PARTICIPATE IN THIS STUDY, WILL IT COST YOU ANYTHING?

There is no cost associated with your participation in this study.

WILL YOU RECEIVE ANY COMPENSATION FOR PARTICIPATING IN THIS STUDY?

Upon completion of this study you will receive $20 either as check or in the form of gift certificate.

WHAT OTHER OPTIONS ARE AVAILABLE IF YOU DO NOT WANT TO TAKE PART IN THIS STUDY?

You understand that your consent to participate in this research is entirely voluntary, and that your refusal to participate will involve no prejudice, penalty or loss of benefits to which you would otherwise be entitled.

CAN YOU WITHDRAW FROM THIS STUDY?

If you consent to participate in this study, you are free to stop your participation in the study at any time without prejudice, penalty, or loss of benefits to which you would otherwise be entitled

HOW WILL THE CONFIDENTIALITY OF YOUR RECORDS BE PROTECTED?

The researcher seeks to maintain the confidentiality of all data and records associated with your participation in this research. There are rare instances (e.g., according to policy, contract or regulation), however, that may arise from capturing photographs where the researcher is required to share personally-identifiable information (e.g., child abuse, threatened violence against self or others). For example, in response to a complaint about the research, officials at the University of New Hampshire, designees of the sponsor(s), and/or regulatory and oversight government agencies may access research data.

All data recorded throughout this study will be stored on Project54 servers which may only be accessed by members of the Project54 team. Textual logs and voice recordings will be made throughout this study for transcriptional purposes regarding the performance of Project54's imaging application. For organizational purposes recorded data will be labeled with user identification numbers. These identification numbers are retrieved from Flickr; therefore, if you truly wish to remain anonymous, you can either disallow access to your account by the imaging application or remove all identifiable information from your account. All data will be disclosed in a generically-labeled anonymous manner. Disclosures include: other researchers inside and outside the University of New Hampshire; conferences; journals; presentations; and all other general publications, internet related and not. These files will be stored for an undetermined amount of time after the closure of this study for the case of future analysis.

WHOM TO CONTACT IF YOU HAVE QUESTIONS ABOUT THIS STUDY

If you have any questions pertaining to the research you can contact Michael A. Farrar at mafarrar@unh.edu or Andrew L. Kun at andrew.kun@unh.edu to discuss them.

If you have questions about your rights as a research subject you can contact Julie Simpson in the UNH Office of Sponsored Research, 603-862-2003 or Julie.simpson@unh.edu to discuss them.

I, ______________________ CONSENT/AGREE to participate in this research study

______________________________ ____________

Signature of Subject Date

APPENDIX B

PRE- AND POST-EXPERIMENT QUESTIONNAIRES

<u>**Using Voice to Tag Digita Photographs on the Spot**</u>

Pre-experiment Questionnaire

Please complete the following survey concerning your experiences using general imaging hardware and software. Along-side each of the statements you will find an area designated for your response, which may take on a value in the range of 1 to 5: 1 being strongly disagree, 2 disagree, 3 undecide, 4 agree and 5 strongly agree. Some portions of the survery may require more detailed information. Please provide your response in the area below such sections. Space has also been provided below each statement for any additional comments you may have. If you feel that a question does not apply to you, then do not respond to it, but please indicate why it does not apply. Thank you.

DATE: ____________________________ **TIME:** ________________________________

PARTICIPANT ID: _________________ **GENDER: Male / Female** **AGE:** _______

UNIVERSITY DEPARTMENT:

____

EDUCATION: Freshman / Sophomore / Junior / Senior / Graduate

Statement	Response				
	Strongly Disagree				Strongly Agree
	1	2	3	4	5
1. I have used digital cameras on multiple occasions.					
2. I own a digital camera. Yes / No.	x	x	x	x	x
3. How many pictures do you take on a monthly basis with a digital camera? Number of pictures: ________	x	x	x	x	x
4. I own a mobile phone. Yes / No. Number of years you've owned a mobile phone: _________ years.	x	x	x	x	x
5. My mobile phone has a full/slideout keyboard. Yes / No.	x	x	x	x	x
6. My mobile phone uses T9 text entry. Yes / No.	x	x	x	x	x
7. I often type on my mobile phone.					

Statement	Response				
	Strongly Disagree				Strongly Agree
	1	2	3	4	5
8. I use a stylus to type on my mobile phone.					
9. I use a full/slideout keyboard to type on my mobile phone.					
10. I often send text messages from my mobile phone.					
11. I own a camera phone. Yes / No.	x	x	x	x	x
12. How many pictures do you take on a monthly basis with a camera phone? Number of pictures: __________.	x	x	x	x	x
13. I'm familiar with desktop/internet photo managers which allow me to tag images.					
14. I tag my pictures.					
15. Tagging pictures is useful.					
16. I currently have an account with a photo-sharing website (e.g. Facebook, Flickr, etc.). Yes / No. Please list websites and year/month you opened the account:	x	x	x	x	x
17. I upload my pictures to the above website(s).					
18. How often do you visit photo-sharing websites? Daily / Weekly / Monthly / Rarely / Never Please list websites:	x	x	x	x	x
19. How many pictures do you view on photo-sharing websites on a monthly basis? Number of pictures: __________.	x	x	x	x	x
20. I view the tags of the pictures from question 19.					
21. I have experience using a PDA, a Smartphone, an iPhone, or a similar device.					
22. My PDA, Smartphone, iPhone, or similar device allows me to upload pictures to a photo-sharing website.					

Statement	Response				
	Strongly Disagree				Strongly Agree
	1	2	3	4	5
23. I have used voice commands to interact with my PDA, Smartphone, iPhone, or similar device. Please give examples (e.g. to dial a number, interact with a search engine, etc.).					
24. I have used speech to interact with computers in the past. Please give examples (e.g. I tried it on my PC, I use it to interact with automated services over the phone, etc.).					

Using Voice to Tag Digital Photographs on the Spot

Post-experiment Questionnaire

Please complete the following survey concerning your experiences using Project54's imaging application. Alongside each of the statements you will find an area designated for your response, which may take on a value in the range of 1 to 5: 1 being strongly disagree, 2 disagree, 3 undecide, 4 agree and 5 strongly agree. Some portions of the survery may require more detailed information. Please provide your response in the area below such sections. Space has also been provided below each statement for any additional comments you may have. If you feel that a question does not apply to you, then do not respond to it, but please indicate why it does not apply. Thank you.

Participant ID: ____________________________Date: __________ Time: _______

Statement	Response: Strongly Disagree 1	2	3	4	Strongly Agree 5
1. I found the capturing of pictures to be simple and intuitive.					
2. I found the tagging of pictures to be simple and intuitive.					
3. I tagged most of the pictures I took during this experiment.					
4. I understand how to tag photographs using voice commands.					
5. I found voice commands to be more convenient than typing when tagging pictures.					
6. I understand the concept of the tag bank.					
7. I found the tag bank to be helpful when reusing previously entered tags.					
8. After the initial setup of the application the tag bank contained most of the tags I needed.					
9. After the initial setup of the application I removed many tags from the tag bank.					
10. It was easy to use the keyboard.					
11. It was easy to type tags.					
12. I used the keyboard to tag most of the pictures.					
13. I used the stylus to tag most of the pictures.					

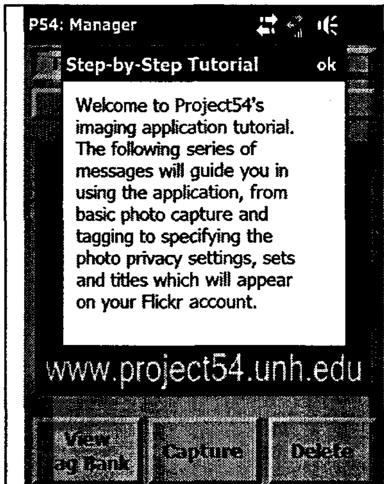
Statement	Response				
	Strongly Disagree				Strongly Agree
	1	2	3	4	5
14. The imaging application responded correctly to my tactile (button) commands.					
15. The imaging application responded correctly to my voice commands.					
16. The "Talk" button was easy to operate.					
17. The imaging application recognized most of my utterances.					
18. Using voice commands was frustrating.					
19. Using voice commands worked well outdoors.					
20. Using voice commands worked well indoors.					
21. Using voice commands worked well noisy environments.					
22. I was satisfied with the quality of the camera.					
23. The GUI buttons and fields were large enough to allow easy interaction with the application.					
24. The tutorial(s) detailed the imaging application in a clear manner.					
25. I had to reference the tutorial(s) many times while using the imaging application.					
26. I enjoyed participating in this experiment.					
27. I liked today's weather.					
28. Today was a difficult day at school and/or work.					
29. Tagging pictures is useful.					
30. Tagging pictures immediately after they are taken is useful.					
31. I would install a free camera phone application that allowed me to tag pictures using voice commands and upload them to a photo sharing website.					
32. I would install a free camera phone application that allowed me to tag pictures using a GUI (but not voice commands) and upload them to a photo sharing website.					

Statement	Response				
	Strongly Disagree				Strongly Agree
	1	2	3	4	5
33. I would pay for a camera phone application that allowed me to tag pictures using voice commands and upload them to a photo sharing website.					
34. I would pay for a camera phone application that allowed me to tag pictures using a GUI (but not voice commands) and upload them to a photo sharing website.					

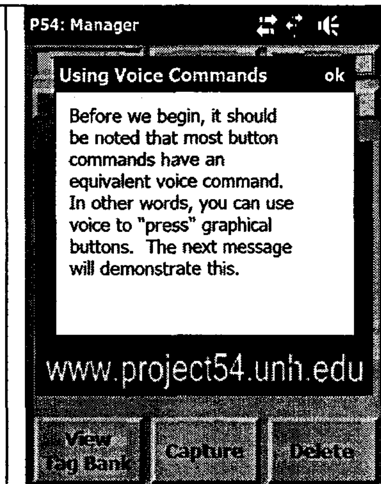
Please use the space below to provide comments and suggestions about the Project54 Imaging Application.

APPENDIX C

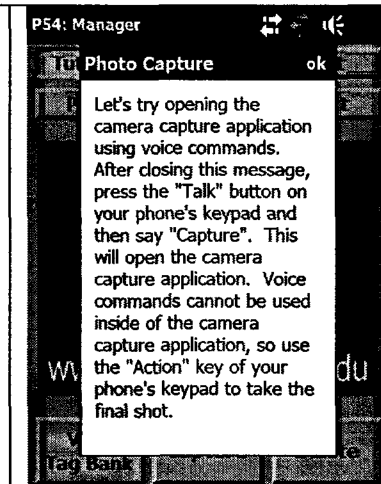
PROMPTS OF THE ON-DEVICE TUTORIAL



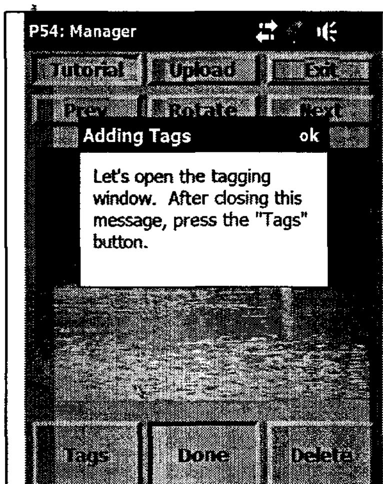
Prompt 1



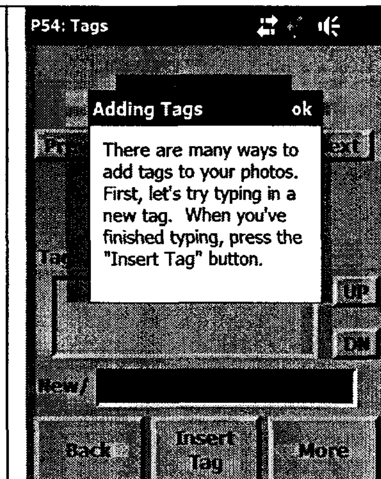
Prompt 2



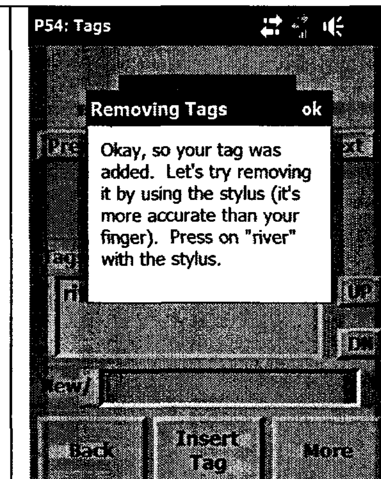
Prompt 3



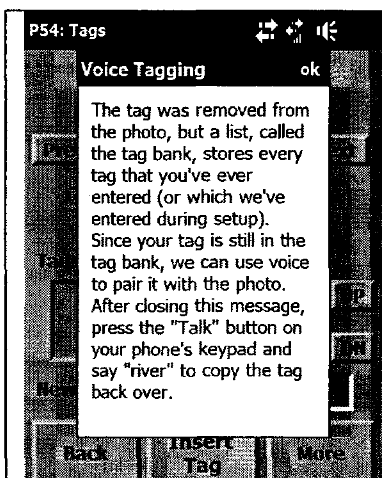
Prompt 4



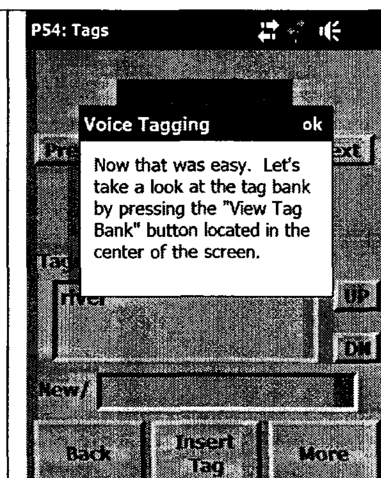
Prompt 5



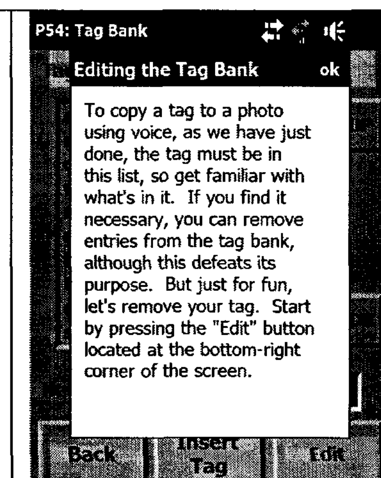
Prompt 6



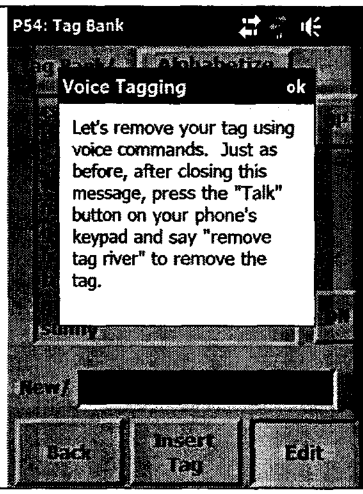
Prompt 7



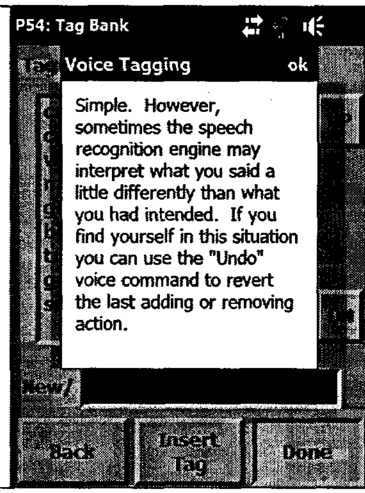
Prompt 8



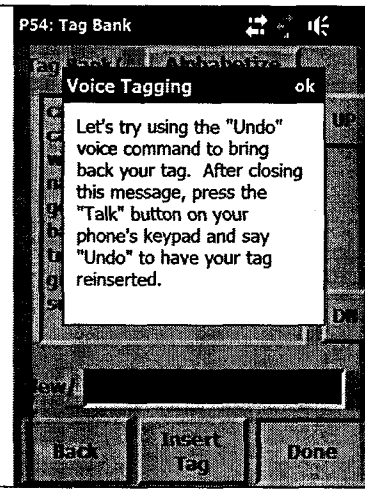
Prompt 9



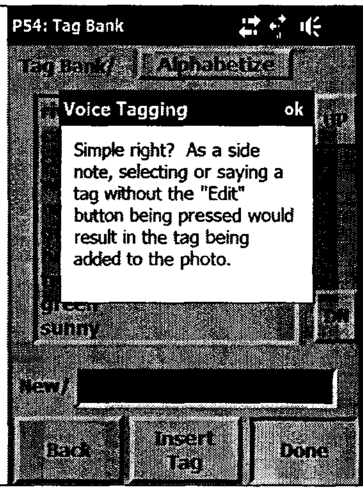
Prompt 10



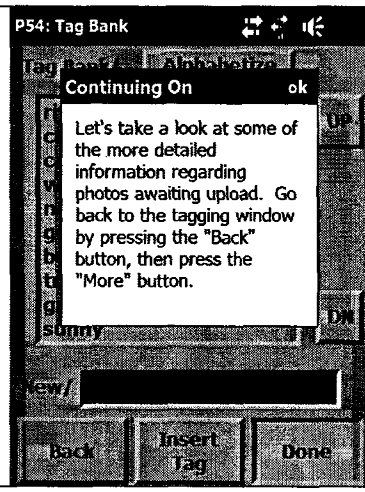
Prompt 11



Prompt 12

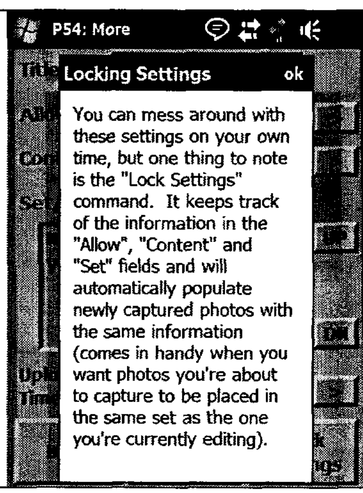


Prompt 13

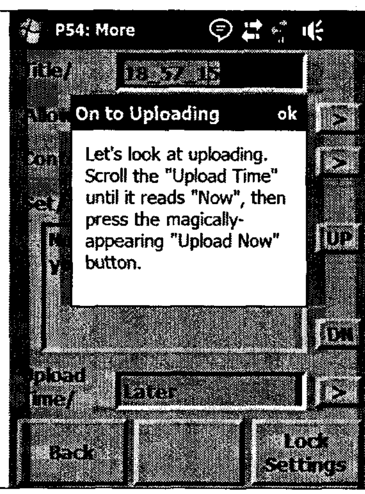


Prompt 14

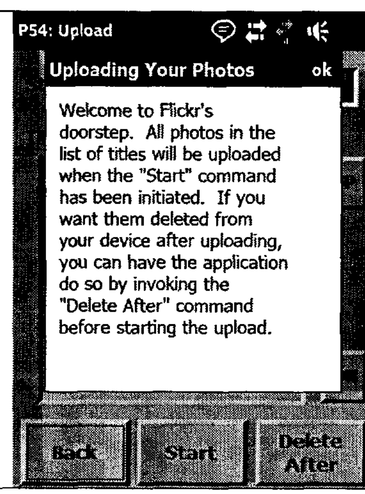
Prompt 15



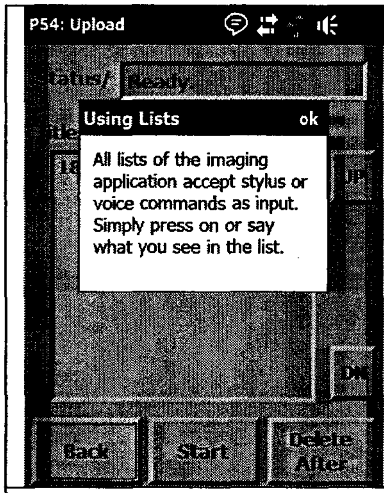
Prompt 16



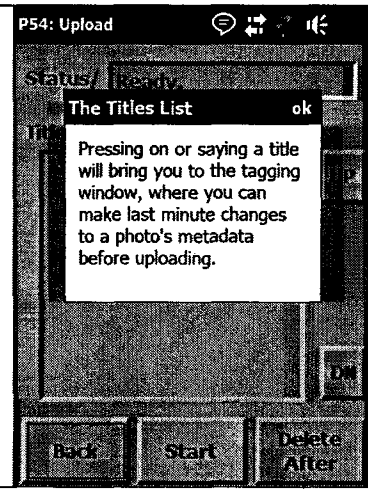
Prompt 17



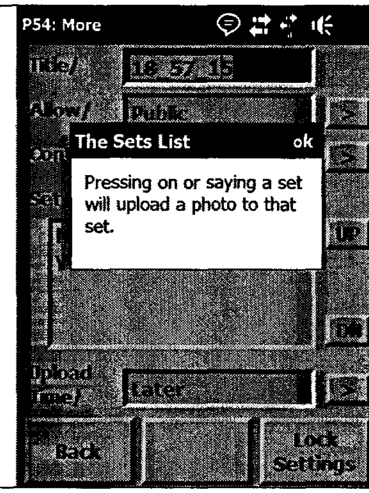
Prompt 18



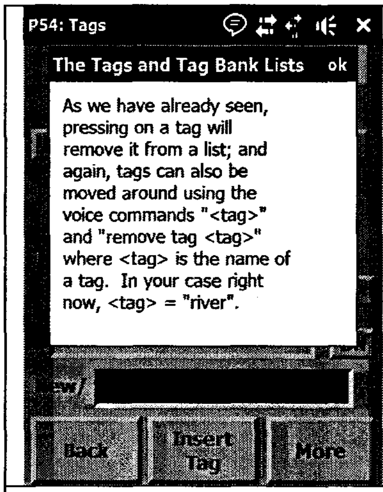
Prompt 19



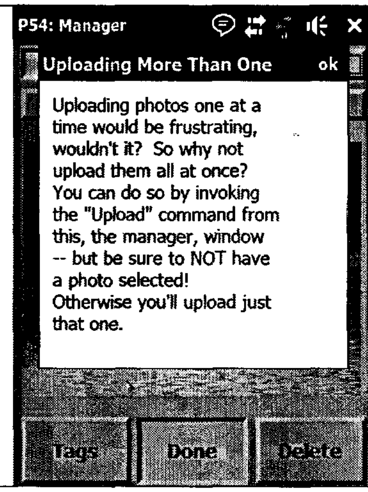
Prompt 20



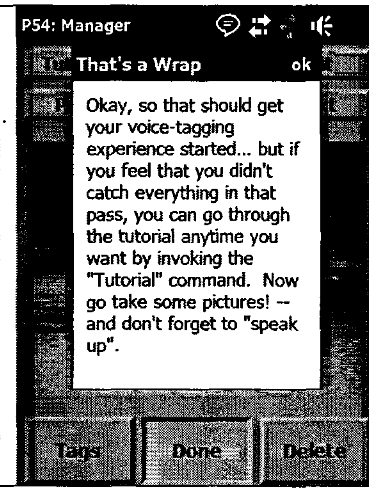
Prompt 21



Prompt 22



Prompt 23



Prompt 24

APPENDIX D

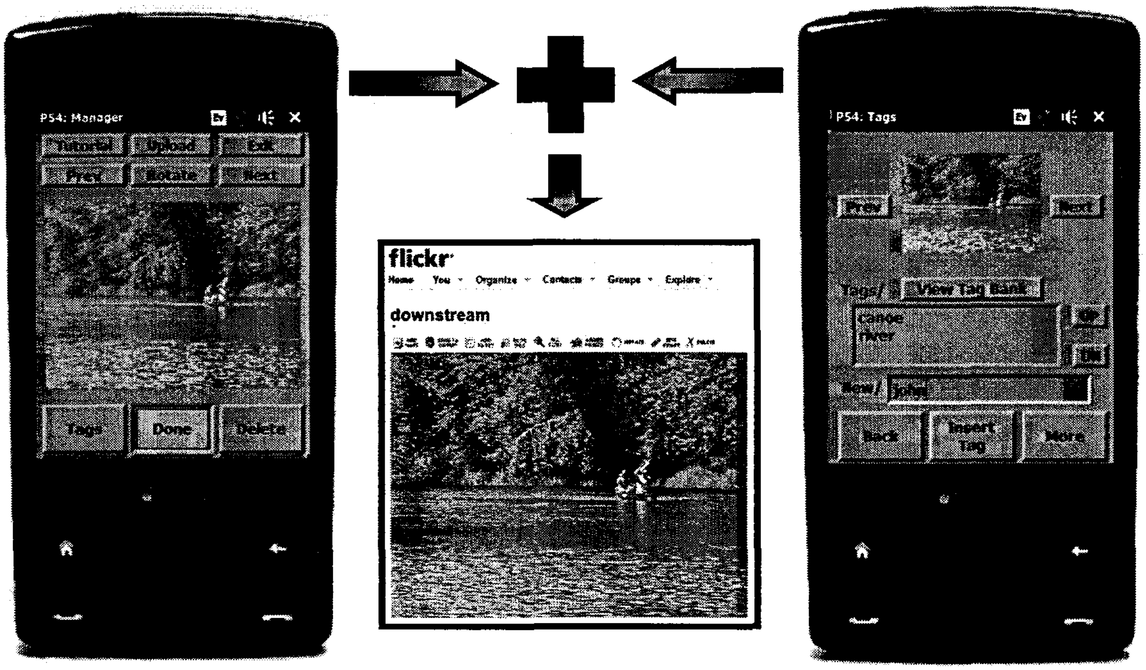
WRITTEN TUTORIAL

Using Voice to Tag Digital Photographs on the Spot

Project54's Imaging Application Tutorial

Conducted by Michael A. Farrar – mafarrar@unh.edu

This tutorial will guide you in the usage of Project54's imaging application. Some statements may only relate to HTC Touch series devices, as this was the design platform; however, the imaging application should support all Windows Mobile 6.0 devices. Please review the tutorial in complete, and thank you for your participation.



➢ ***Battery, Reset, and Loading*:**

- The ON/OFF button is the rectangular button located on the top of the device. ***The imaging application will run slower when the device is first turned on. Please be patient, things will speed up.***
- If the battery dies, the device must be recharged. At next use, the operating system will need to be reconfigured. Follow the on-screen instructions. ***All photographs and their tagging data will be lost. Please keep the device charged to avoid losing this data.*** Once configured, the imaging application will need to be reinstalled.
- If the device no longer responds to your actions, it can be reset by using the stylus to depress the reset button. The reset button is located either on the bottom or on the backside of the device underneath the battery cover. The cover may be removed by sliding it upwards. ***Please contact me if you find yourself resetting the device often.***
- To use the imaging application you must manually load it by selecting "Start→Programs→Project54" from the upper-left corner of the desktop window.

➢ *Manipulating the Imaging Application*:

The imaging application consists of six windows and an external utility: the First Use window (figure 1), the Manager window (figure 2a), the external camera utility (figure 2b), the Tags window (figure 3), the Tag Bank window (figure 4), the More window (figure 5) and the Upload window (figure 6). The device is touch-sensitive, so you may use the stylus for screen-tap interactions upon the graphical buttons and text listings. Most graphical buttons are also accessible via speech commands (see tables 1-6 for a complete list of speech commands). ***To activate speech recognition, PRESS and RELEASE the talk button (highlighted in red throughout this tutorial) located at the lower-left position of the device's external keypad. Once active, you will have three seconds to say a command.*** Please review the figures and tables of the next sections of this tutorial for complete details of the imaging application's commands and features.

➢ *The First Use Window: Linking with your Flickr account.*

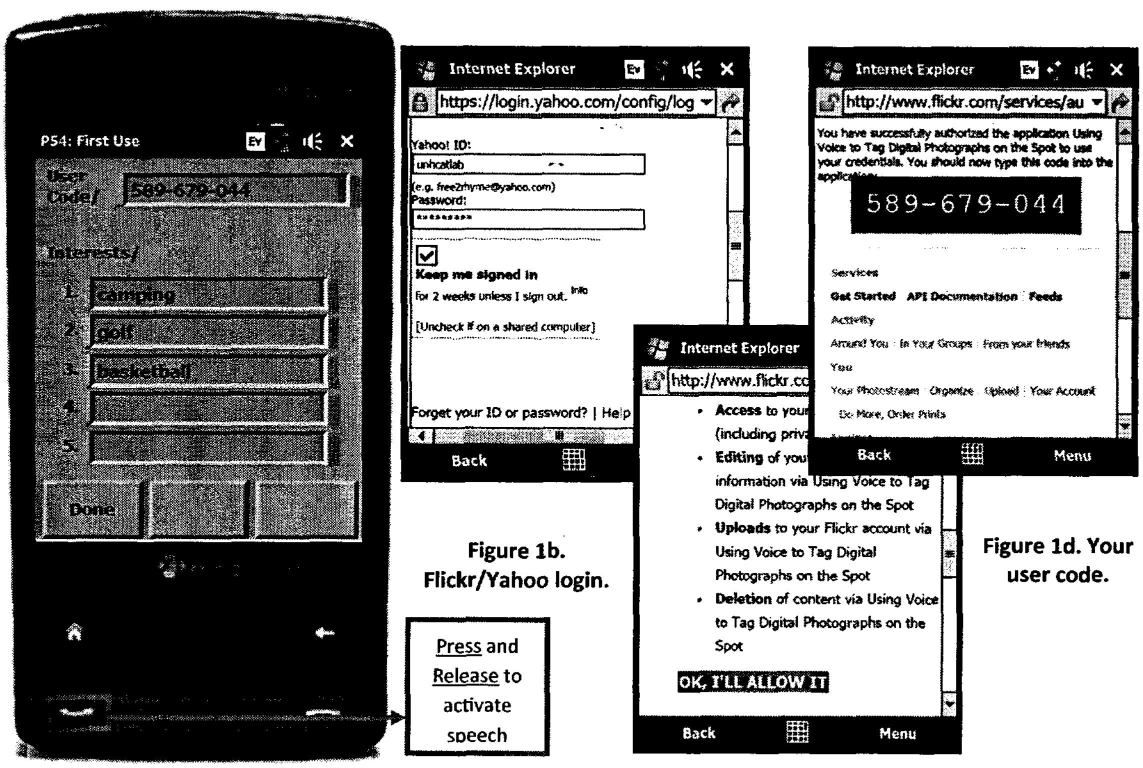


Figure 1b. Flickr/Yahoo login.

Figure 1d. Your user code.

Figure 1a. The First Use window.

Figure 1c. Consent to access your account.

Graphical Button	*Speech Command*	*Function*
Done	"DONE"	Exits the First Use window.

Table 1: Graphical and speech commands of the First Use window.

➤ ***Linking with Your Flickr Account***: Project54's imaging application provides seamless interoperability with Flickr, a popular photo-sharing website. In order to use Flickr's services you must complete Flickr's consent procedure by retrieving your user code. The First Use window will automatically direct you to Flickr's consent URL. The retrieval process is depicted above.

➤ ***Initializing the Tag Bank***: The tag bank is a list of tags which the imaging application stores for your convenience. Use of the tag bank will be detailed in the following sections of this tutorial. During the verification process the application will copy your most frequently used tags from your Flickr account and place them into the tag bank for later use. As an additional effort to expand the initial contents of the tag bank, the imaging application allows for you to enter up to five photographic interests. Each of these will then be compared against Flickr's immense tag database and all matching tags will be placed into the tag bank for later use.

➤ ***The Manager Window***: *Capturing and manipulating photographs.*

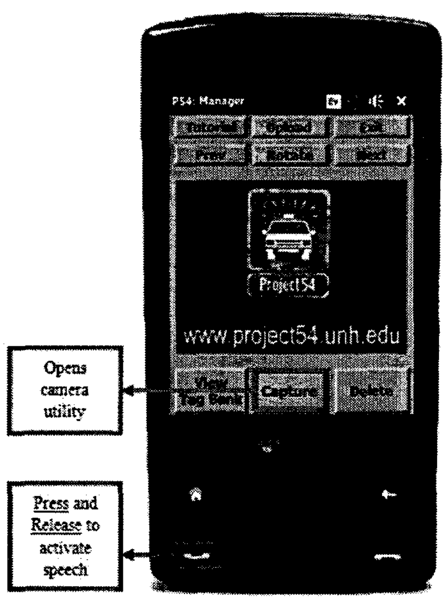


Figure 2a. The Manager window.

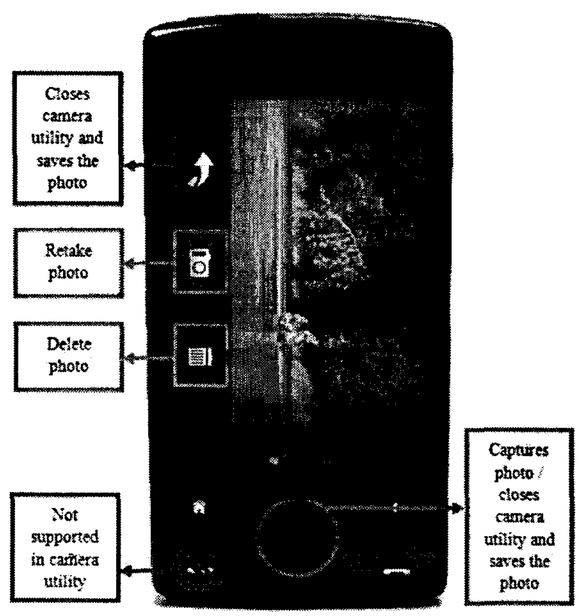


Figure 2b. The external camera utility.

Graphical Button	*Speech Command*	*Function*
Tutorial		Opens the on-device tutorial, guiding your use of Project54's imaging application.
Upload	"UPLOAD"	Uploads specified photograph(s) to Flickr.
Exit		Exits the imaging application.
Prev	"PREV" or "PREVIOUS"	Scrolls to a previously captured photograph in a later-date manner.

Table 2: Graphical and speech commands of the Manager window.

Graphical Button	*Speech Command*	*Function*
Rotate	"ROTATE"	Rotates the current photograph counterclockwise.
Next	"NEXT"	Scrolls to a previously captured photograph in an earlier-date manner.
View Tag Bank / Tags	"VIEW TAG BANK" / "TAGS"	Opens the Tag Bank or Tags window, where you may pair tags with photos.
Capture / Done	***"CAPTURE" / "DONE"***	***Opens / Closes the camera utility for capturing new photos.***
Delete		Deletes the current photograph.

Table 2 (continued): Graphical and speech commands of the Manager window.

Note: The "Action" device-button is used to capture the photograph when the external camera utility is active. The action button is the circular button highlighted in blue. Speech commands are not supported in the utility.

- ***The Tags Window**: Tagging your photographs.*

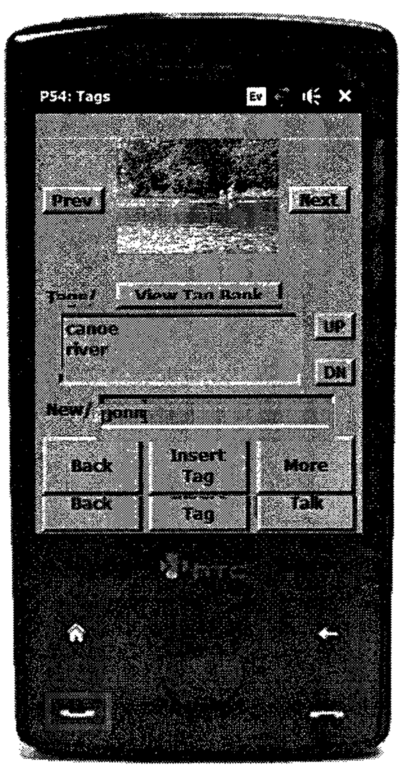


Figure 3. The Tags window.

Graphical Button	*Speech Command*	*Function*
Prev	"PREV" or "PREVIOUS"	Scrolls to a previously captured photograph in a later-date manner.
Next	"NEXT"	Scrolls to a previously captured photograph in an earlier-date manner.
View Tag Bank	"VIEW TAG BANK"	Opens the Tag Bank window, where you can pair tags with a photo or edit the tag bank.
UP	"SCROLL UP"	Scrolls the tags listing in an upward manner.
DN	"SCROLL DOWN"	Scrolls the tags listing in a downward manner.
Back	"BACK"	Returns to the Manager or Upload window.

Table 3a: Graphical and speech commands of the Tags window.

Graphical Button	*Speech Command*	*Function*
Insert Tag	"INSERT TAG"	Pairs a newly typed tag with the photo and stores it in the tag bank.
More	"MORE"	Opens the More window, where you can specify Flickr upload settings.

Table 3a (continued): Graphical and speech commands of the Tags window.

- ***Inserting Tags via Text Entry***: New tags may be ***typed*** into the new tag text field as desired ***and inserted using the insert tag command. Newly typed tags will be paired with the photograph and stored in the tag bank.***

- ***Removing Tags via Touch***: Tags may be ***removed*** from the photo ***by selecting them with the stylus.*** If an incorrect selection has been made, the tag may be reinserted manually or by using the "UNDO" speech command detailed in table 3b.

- ***Removing Tags via Speech***: Tags may be ***removed*** from a photo ***by using the tagging specific speech commands*** of the Tags window. These commands are generalized in table 3b below.

Speech Command	*Function*
"REMOVE TAG <tag_name>"	***Removes the tag specified by <tag_name> from the photograph.*** *For example, saying "REMOVE TAG RIVER" will remove "river" from the photo of figure 3.*
"UNDO"	Removes the lastly added or reassigns the lastly removed tag to the photograph. Following the example above, the undo command would reassign "river" to the photo.

Table 3b: Tagging specific speech commands of the Tags window.

- ***Adding Tags via Speech***: ***Tags may be added to a photograph from the tag bank by using the tagging specific speech commands of the Tag Bank window. These commands are generalized in table 4b. It should be noted that a tag must appear in the tag bank before speech may be used to copy it to the photo. Please review the next section of this tutorial for further detail on adding tags via speech.***

➢ ***The Tag Bank Window***: *Tagging your photographs and storing tags.*

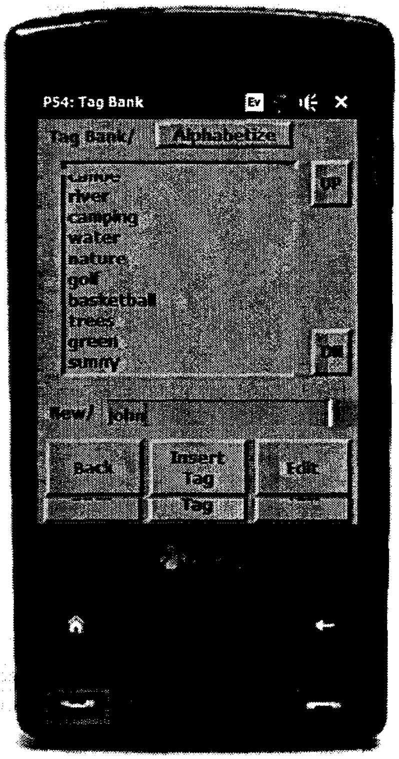


Figure 4. The Tag Bank window.

Graphical Button	*Speech Command*	*Function*
Alphabetize / Sort by Uses	"ALPHABETIZE" / "SORT BY USES"	Orders the tag bank alphabetically or by the tags' frequency of use.
UP	"SCROLL UP"	Scrolls the tag bank listing in an upward manner.
DN	"SCROLL DOWN"	Scrolls the tag bank listing in a downward manner.
Back	"BACK"	Returns to the Manager or Tags window.
Insert Tag	"INSERT TAG"	Pairs a newly typed tag with the photo and stores it in the tag bank.
Edit / Done	"EDIT" / "DONE"	Enters / exits the tag bank editing mode of the Tag Bank window.

Table 4a: Graphical and speech commands of the Tag Bank window.

➢ ***Inserting Tags via Text Entry***: New tags may be ***typed*** into the new tag text field as desired ***and inserted using the insert tag command. Newly typed tags will be paired with the photograph and stored in the tag bank.***

➢ ***Utilizing the Tag Bank via Touch***: Tags may be ***added*** to a photo ***by selecting them with the stylus.*** A notification will be displayed upon adding a tag. If an incorrect selection has been made, the tag may be removed manually or by using the "UNDO" speech command detailed in table 4b.

➢ ***Utilizing the Tag Bank via Speech***: Tags may be ***added*** to a photo ***by using the tagging specific speech commands*** of the Tag Bank window. These commands are generalized in table 4b below.

Speech Command	*Function*
"<tag_name>"	***Adds the tag specified by <tag_name> to the photograph. For example, saying "RIVER" will add "river" to the photo of figure 3.***
"UNDO"	Removes the lastly added or reassigns the lastly removed tag to the photograph. Following the example above, the undo command would remove "river" from the photo.

Table 4b: Tagging specific speech commands of the Tag Bank window.

Note: The Tag Bank window does not need to be visible in order to tag photographs using speech. If a tag is known to exist in the tag bank, it may be added to the photo from within the Tags window using the commands defined above.

- ***Editing the Tag Bank**: Tags may be removed from the tag bank by selecting them with the stylus or by using the tagging specific speech commands of the Tag Bank window under its tag bank editing mode. These commands conform to the same rules defined above in table 3b for removing tags.*

- ***The More Window**: Specifying Flickr upload settings.*

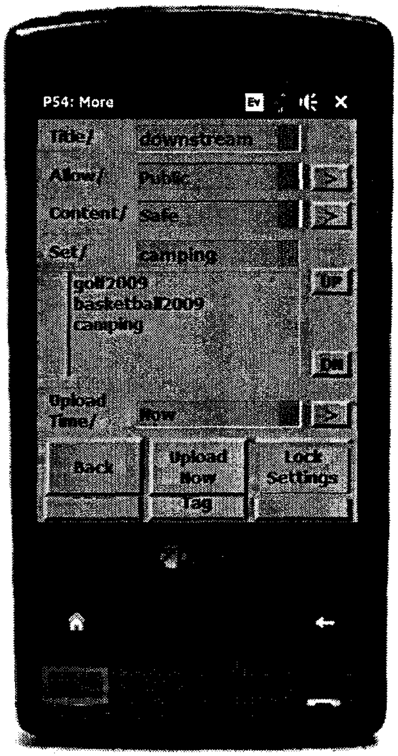


Figure 5. The More window.

Graphical Button	*Speech Command*	*Function*
Allow >	"NEXT ALLOW"	Scrolls the photo's viewing audience.
Content >	"NEXT CONENT"	Scrolls the photo's content type.
UP	"SCROLL UP"	Scrolls the sets listing in an upward manner.
DN	"SCROLL DOWN"	Scrolls the sets listing in a downward manner.
Upload Time >	"NEXT UPLOAD TIME"	Scrolls the photo's upload time.
Back	"BACK"	Returns to the Tags window.
Lock Settings / Unlock Settings	"LOCK SETTINGS" / "UNLOCK SETTINGS"	Saves the current configuration of the More window.

Table 5a: Graphical and speech commands of the More window.

Graphical Button	*Speech Command*	*Function*
Upload Now	"UPLOAD NOW"	Opens the Upload window for immediate photo upload.

Table 5a: Graphical and speech commands of the More window.

- ***Specifying Flickr Upload Settings***: The imaging application allows for you to specify four upload properties relative to a photo's appearance on your Flickr account: a photo's ***title***, ***viewing audience***,

content type and ***set***. These settings may be saved using the ***lock settings*** command, writing the same viewing audience, content type, set and upload time values to newly captured photographs.

- *Title*: The photo's timestamp will be used as its initial title. In figure 5 the title has been changed to "downstream".
- *Viewing Audience*: The viewing audience restricts certain groups of people from viewing your photographs on Flickr. Flickr allows for five viewing audiences: Public; Family + Friends; Family; Friends; Just Me.
- *Content type*: The content type defines the nature of the photograph. Flickr allows for three content types: Safe; Moderate; Restricted.
- *Set*: The imaging application will automatically download your Flickr photosets each time you access the More window. A set may be ***selected*** from among them ***by using the stylus or by using the sets specific speech commands of table 5b***. New sets may be ***created by typing a set name in to the set text field***.
- *Upload Time*: The imaging application allows you to upload photographs at times of most convenience. Upload times consist of: Later; Never; Now.

➢ ***Specifying an Upload Set***: Sets may be ***selected*** for photo upload ***by pressing on them with the stylus or by using the sets specific speech commands*** of the More window. These commands are generalized in table 5b below.

Speech Command	*Function*
"<set_name>"	***Adds the photo to the set specified by <set_name>. For example, saying "CAMPING" will upload the photo to the "camping" photoset on your Flickr account, as shown in figure 5.***

Table 5b: Sets specific speech commands of the More window.

➢ ***The Upload Window***: *Uploading your photos to Flickr.*

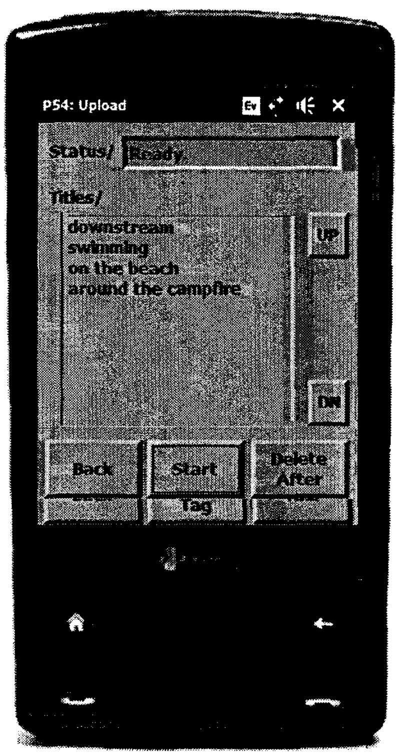


Figure 6. The Upload window.

Graphical Button	*Speech Command*	*Function*
UP	"SCROLL UP"	Scrolls the titles listing in an upward manner.
DN	"SCROLL DOWN"	Scrolls the titles listing in a downward manner.
Back	"BACK"	Returns to the Manager or More window.
Start	"START"	Initiates photo upload.
Delete After	"DELETE AFTER"	Deletes photographs after uploading.

Table 6a: Graphical and speech commands of the Upload window.

➢ ***Making Last-Minute Changes to Your Photographs***: Titles may be ***selected*** for editing ***by pressing on them with the stylus or by using the titles specific speech commands*** of the Upload window. These commands are generalized in table 6b below.

Speech Command	*Function*
"<title_name>"	***Selects the photo specified by <title_name> for editing. For example, saying "DOWNSTREAM" will open the Tags window with the photograph entitled "downstream", as shown in figure 6.***

Table 6b: Titles specific speech commands of the Upload window.

Note: Entering the Upload window from the Manager window without a photograph selected/viewed will populate the upload listing with all photographs pending upload. Uploading multiple photographs saves time.

APPENDIX E

INSTITUTIONAL REVIEW BOARD APPROVAL

University *of* New Hampshire

Research Integrity Services, Office of Sponsored Research
Service Building, 51 College Road, Durham, NH 03824-3585
Fax: 603-862-3564

30-Mar-2009

Farrar, Michael A.
Elec/Computer Engineering, Kingsbury
10 Brickyard Lane
Nashua, NH 03062

IRB #: *4517*
Study: Using Voice to Tag Digital Photographs on the Spot
Approval Date: 30-Mar-2009

The Institutional Review Board for the Protection of Human Subjects in Research (IRB) has reviewed and approved the protocol for your study as Expedited as described in Title 45, Code of Federal Regulations (CFR), Part 46, Subsection 110.

Approval is granted to conduct your study as described in your protocol for one year from the approval date above. At the end of the approval period, you will be asked to submit a report with regard to the involvement of human subjects in this study. If your study is still active, you may request an extension of IRB approval.

Researchers who conduct studies involving human subjects have responsibilities as outlined in the attached document, *Responsibilities of Directors of Research Studies Involving Human Subjects.* (This document is also available at http://www.unh.edu/osr/compliance/irb.html.) Please read this document carefully before commencing your work involving human subjects.

If you have questions or concerns about your study or this approval, please feel free to contact me at 603-862-2003 or Julie.simpson@unh.edu. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson
Manager

cc: File
Kun, Andrew