Summer 2015

Effects of Input Modality on Capturing Notes

Chaya Bijani
San Jose State University

Follow this and additional works at: https://scholarworks.sjsu.edu/etd_theses

EFFECTS OF INPUT MODALITY ON CAPTURING NOTES

A Thesis

Presented to

The Faculty of the Graduate Program in Human Factors and Ergonomics

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Chaya Bijani

August 2015

The Designated Thesis Committee Approves the Thesis Titled

EFFECTS OF INPUT MODALITY ON CAPTURING NOTES

by

Chaya Bijani

APPROVED FOR THE GRADUATE PROGRAM IN HUMAN FACTORS AND
ERGONOMICS

SAN JOSÉ STATE UNIVERSITY

August 2015

Dr. Sean Laraway	Department of Psychology
Dr. Emily Wughalter	Department of Kinesiology
Brent-Kann White	ChargePoint® Inc.

ABSTRACT

EFFECTS OF INPUT MODALITY ON CAPTURING NOTES

by Chaya Bijani

The features of the smartphone make it an indispensible commodity of Western urban lifestyles. However, the most common problems of using a mobile device for work-related activities are limited screen space and poor input techniques. People in the workforce whose daily job entails being in a mobile environment generally prefer to carry light, mobile devices along with a pen and a notepad. The purpose of this study was to investigate optimal input modality for taking notes. The three modes of input evaluated were spoken notes, typing on the phone, and writing by hand using a pen and paper. The variables measured to evaluate the three modalities were accuracy of content, perceived mental task load, preferred mode, and number of words. Spoken notes were significantly more accurate, less taxing mentally, and more detailed compared to typed or handwritten notes. The difference between typed and handwritten notes was shown to be non-significant. However, the majority of participants preferred the typed or handwritten modality. The study shows that even though the accuracy of the spoken modality by far exceeded the rest, spoken notes are best suited for taking rough notes for personal use only.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to committee members Dr. Sean Laraway, Dr. Emily Wughalter and Brent White for supporting and encouraging me throughout my thesis process. To Dr. Laraway, thank you for being patient and guiding me throughout the thesis process by keeping me focused. To Dr. Wughalter, I would not have finished my project without your invaluable emotional and academic guidance during the final phase of thesis work. Thank you for being there for me. And to Brent, thank you for extending unwavering support to the thesis project from its inception to its conclusion. I learned so much from you.

I would like to thank my family for believing in me and encouraging me through my entire journey in the master's program. To my dearest sister, Asha: Thank you for your enduring support and encouragement, enforcing every single day that I can achieve this. I cannot thank you enough for helping me transcribe the data and proofreading the document untold times.

Finally, I am thankful to have friends like John Cartan, Shobana Subramanian-Arora and Gül Yayli. To John, thank you for spending countless hours listening and offering inspiring ideas to guide the initial direction of my thesis. Shobana, you were my sounding board. Thank you for always being available to discuss my thesis. And to Gül, thank you for inviting me over, without fail, every single day for delicious Turkish coffee when I was mired in my thesis and needed a break.

To my family, friends and professors, thank you! I shall remain forever in your debt.

TABLE OF CONTENTS

List of Figures

List of Tables

Introduction

The ability to remain constantly connected has made the smartphone the most-used mobile technology in the world. Since the advent of touch screen smartphones in 2007, a multitude of smartphone devices have flooded the market and their popularity has exploded throughout the world (Rivera & Van Der Meulen, 2014). Apart from making phone calls, a smartphone is a mode of entertainment (playing games, watching videos), a way to remain socially connected (checking/responding to emails), a navigational aid (GPS), and a handy tool to find information (searching for restaurants). Most people these days would not think about leaving their homes without these handheld devices that function as mini computers.

Despite the convenience of using a smartphone for these activities, when using smartphones for work the most common problems are limited screen space and poor input techniques. Because of visual display limitations, interacting with smartphones, especially on the go, places heavy demands on attentional resources and physical capabilities of mobile users (Tamminen, Oulasvirta, ToisKallio, & Kankainen, 2004). Empirical studies corroborate the inconvenience of reading and typing on small, mobile devices. For example, Hoggan, Brewster and Johnston (2008) found that typing on small screens is ergonomically inconvenient. The small size of the icons and buttons leads to task errors and increases time spent on tasks (Parhi, Karlson, & Bederson, 2006). These results are in line with Fitts' Law (1954), which established that target size is inversely proportional to the time it takes to hit that object. Fitts' Law was expanded and reevaluated for use on a touch-screen handheld device and the results of this experiment

confirmed the original findings (Bi, Li, & Zhai, 2013).

Motivation

Because of the popularity of smartphones, enterprise application vendors offer

mobile solutions for their desktop/laptop applications so that mobile users can easily

access information outside of the office (Rampoldi-Hnilo, White, Snyder, & Sampanes,

2009). A common job for people working outside the office is a field sales representative

(sales rep). Sales reps are frequently in the field, and they prefer to carry light, portable

mobile devices. One of the requirements of sales reps is to document sales activity so that

they can accurately forecast future sales in order to assess their performance. Sales reps

mostly compile notes in the parking lot or in their cars right after meetings; some sales

reps type notes on their smartphones, others use a pen and notepad to take handwritten

notes and some use applications like "Voice Memos" or "Evernote" on their mobile

phone to record verbal notes (Bijani, White, & Vilrokx, 2013). The research question of

interest was which interface would allow sales reps to document their sales activity

efficiently and quickly.

Recall from Working Memory

Humans constantly reference information stored in the brain to act upon current

tasks. Information from recent events lives in the temporary storage of the human brain

known as working memory. According to the working memory model proposed by

Baddeley & Hitch (1974) and Baddeley (2000), working memory is made up of three

components: the phonological loop, the visuo-spatial sketchpad and the central executive

system. The phonological loop processes auditory input. Auditory information quickly

decays unless continuously rehearsed in the phonological loop. The visuo-spatial sketchpad is responsible for processing mental images. The central executive system manages information from these two sub-systems to perform cognitive tasks. During recall, the central executive system employs working memory to process information to produce coherent information (Baddeley A., 2000).

Process of Writing

The writing process applies problem-solving strategies to organize and structure content to be written (Flower & Hayes, 1981). Flower and Hayes (1981) posited writing to be a complex cognitive process that involves planning, translating, reviewing and monitoring information to be written. Based on the cognitive process theory of writing proposed by Flower and Hayes (1981) and model of working memory proposed by Baddeley (2000), Kellogg (1996) proposed a writing-process model that describes the engagement of working memory in producing written material. The writing process engages working memory to organize and structure the details to be written. The visuo-spatial sketchpad is used to plan, organize and visualize content and the phonological loop is employed for translating content. Chenoweth and Hayes (2003) conducted an experiment to validate the role of the phonological loop in writing. Participants described multiple cartoon strips under different conditions. In one condition, participants repeated a syllable simultaneously while writing to disrupt the rehearsal process of the phonological loop. The experiment concluded that the secondary task interfered with the articulatory rehearsal process, resulting in shorter written sentences and more errors. Kellogg, Olive, & Piolat (2007) verified the engagement of both sub-components of

working memory in producing longhand written answers. During the experiment, participants wrote definitions of nouns while performing a parallel task. The parallel task required participants to identify a syllable, a shape or location of the stimulus that matched the recently presented stimulus. Participants took longer to complete the writing assignment in the presence of the interference that tampered with the information present in the phonological and visual components of the working memory. The findings prove that both the phonological loop and the visuo-spatial sketchpad are used during the writing process.

In the current study, writing transcription is evaluated under two conditions: typed and handwritten. Based on the information above, typed or handwritten notes utilize working memory to store information of recent events as well as to organize sentences to produce written content. Given that human working memory can only hold 3-5 chunks of information at any given time (Cowan, 2001), the accuracy of typed or handwritten notes may suffer as multiple processes compete for limited working memory resources.

Process of Speaking

Speech, on the other hand, is an inherently human trait and the human brain is uniquely equipped for speech (Nass & Gong, 2000). Sound perception begins soon after birth, and language learning follows suit. A three- year-old child is equipped to comprehend complex language syntax and grammatical formations (Lieberman, 1993), long before that child learns to write or type. Parts of the brain, namely Broca's area and Wernicke's area, are dedicated to processing and producing speech. The central executive system engages Broca's area, Wernicke's area and the phonological loop to process, plan

and organize sentences before words are actually made audible (Jacquemot & Scott, 2006). This implies that the visuo-spatial sketchpad, a sub-component of working memory, is not utilized in speech production. Based on the Multiple Resource Theory (Wickens, 2002; 2008), when two tasks use different resources, time-sharing demands on information processing are efficient and there is no cognitive overload. Further written transcription involves several motor sequences to be carried out to achieve the orthographic output. The transcription process also interferes with word storage, resulting in loss of information from working memory (Bourdin & Fayol, 2000). Overloading working memory creates a bottleneck for information processing as multiple demands are made on sharing the same resource (Wickens, 2002; 2008). The literature discussed thus far suggests that spoken notes may be more accurate than typed or handwritten notes. It is also inferred that spoken notes might place less demand on working memory as it engages other parts of the brain to complete the task. During the speech process the visuo-spatial sketchpad, a sub-component of working memory, is freed up and might aid in retaining more information or visual cues from recent events. In the case of typed or handwritten notes, it is possible that less information from recent events will be transformed into words on paper or on electronic media as working memory has to free up space to plan, organize and structure content to be written.

Another thing that might impact the accuracy of typed and handwritten content is the knowledge of results. Knowledge of results is a verbal or augmented feedback provided at the end of the task to inform the performer about the quality of the task (Salmoni, Schmidt, & Walter, 1984). In a motor-learning paradigm, increasing the

frequency of verbal feedback in between the trials curtails performance (Winstein & Schmidt, 1990). In the current scenario, visual feedback will be constantly available while typing or writing by hand. Lyons, Plaisted and Starner (2004) conducted an experiment to investigate typing speed and accuracy on mobile devices under two conditions. In the first condition, typed feedback was visible on the screen while in the second condition, visual feedback was hidden and only the cursor movement indicated the progression of typed words. The latter condition resulted in fewer errors and improved typing speed. It was concluded that seeing immediate visual feedback was a source of distraction and might have disrupted the flow of information.

Speaking is faster than writing or typing (Basapur, Xu, Ahlenius, & Lee, 2007). Speaking is learned implicitly, whereas learning to type or write by hand is an explicit process. Explicit learning occurs when detailed verbal feedback is given to explain how to perform a task. Children practice penmanship in early years of schooling under constant verbal and visual instruction; this type of learning is an explicit process. Learning to speak is an example of an implicit process; as a child picks up language by listening to others speak. No specific instructions explaining how to move vocal chords to produce sound is given to a child; they learn to do so implicitly or naturally. The knowledge structures formed in the brain from implicit processes are different than those formed by explicit processes (Dienes & Perner, 1999). Implicit processes are faster while explicit processes are comparatively slower as they are sequential and make use of working memory to carry out a task (Maxwell, Masters, Kerr, & Weedon, 2001). From this literature it is gathered that spoken notes can be done faster and impose less cognitive

load. To recap the points established thus far, spoken notes might be more accurate than typed or handwritten notes and they also incur less cognitive load while comprised of more words and sentences.

Typing versus Writing by Hand

Typing has the advantage of having letter keys displayed in the form of a QWERTY keyboard, and the brain uses both recognition and repetition to identify the character to tap; whereas in case of writing by hand the characters are recalled from the long-term memory store and manually transferred on paper. Recall is a two-step process. First, the character is fetched from the memory store and second, the familiarity process kicks in to recognize the character (Kintsch, 1970). Writing by hand employs the two-step recall process to write each character whereas typing employs a one-step recognition process to identify and tap the character on the keyboard. Writing by hand is much more involved than typing on a touch screen, as it needs more cognitive resources to produce the final output. Complex cognitive processes need working memory resources to complete the task and secondary tasks are generally compromised (Wickens, Multiple resources and mental workload, 2008). The secondary task in this scenario will be information from recent events. With that in mind, it is inferred that typing might result in better accuracy of content and might be less strenuous than taking handwritten notes.

Empirical research shows that touch-screen typing ranges from 20-30 words per minute (Sears, Revis, Swatski, Crittenden, & Shneiderman, 1993). Gould, Greene, Boies, Meluson and Rasamny (1990) established typing speed using a soft keypad to be in the range of 30 words per minute. Handwriting speeds are estimated in the range of 10-22

words per minute (Newell & Card, 1985). Given the respective ranges, it is expected that the handwritten mode might result in the least number of words as compared to the other two modalities. Number of words is measured in this study to validate the findings from previous studies as well as to observe whether the difference in number of words impacts accuracy of content.

Current Study

The purpose of this paper was to examine the effects of input modality while taking notes. Three input modalities -- spoken, typed and handwritten -- were evaluated for taking notes. Measures used to investigate the modality were accuracy of content, perceived mental load in using a modality, the user-preferred mode and number of words used while capturing notes of recent events. The findings of this study will augment the existing empirical knowledge and help designers create efficient input techniques for mobile devices.

Research Questions

A range of hypotheses was looked into to evaluate notes captured using various modalities. The following dependent measures were studied in this research:

Accuracy of the content - This variable informed which modality resulted in accurate notes.

Mental Task Load - Cognitive load experienced while using a particular modality was measured by using NASA Task Load Index (NASA-TLX). NASA-TLX has been widely used in mobile studies to capture self-reported mental stress (Barnard et al., 2005; Price et al., 2006).

Preferred Mode - This variable documented participants preferred mode for taking notes.

Number of words - The word count informed the amount of details captured in notes.

Based on the literature discussed above, it was hypothesized that the spoken mode would generate the most accurate notes among the three modalities and typed notes would be more accurate than handwritten notes. The spoken mode might also result in the least mental load, and typing would be less strenuous than writing by hand. Furthermore, it was anticipated that spoken notes would be more detailed, resulting in higher word count. Typed notes would be more detailed than to handwritten notes.

Method

Design

To analyze the data, a repeated-measure one-way analysis of variance was employed with post-hoc tests and Cohen's d as an effect size measure for the comparisons between each modality. The independent variable was input modality with three levels: spoken, typed and handwritten.

The stimuli used in the experiments were three TED Talks videos on general topics. The length of each video was approximately 3 minutes. After browsing through multiple TED talks, these three videos were selected - "Why is 'x' the unknown?" (Moore, 2012), "8 secrets of success" (St. John, 2005), and "Teach statistics before calculus!" (Benjamin, 2009). Moore (2012) talked about how the unknown expression represented by the letter "x" came to be. The speaker traced the origin of letter "x" to Arabic literature and talked about the issues associated in translating Arabic into Latin. St. John (2005) summarized eight keys to success in his talk. He preached concepts such as passion, persistence, ideas and getting pushed as the main contributors of success. Benjamin (2009) outlined how the current high school mathematics curriculum is outdated because it focused on calculus as the summit of mathematical learning. Instead, the speaker stated that statistics should be the fundamental aspiration of mathematics instruction because of its usefulness and relevance in the digital world. All the three videos were carefully selected so that participants were not required to have additional domain knowledge to understand the content. The content of each video was assumed to be of equal complexity. After the videos were selected, three master summaries were

written, one for each video, highlighting the main points of the talk. Master summaries were not the exact transcriptions of the video, but rather a comprehensive summary that conveyed the essence of the video. Based on these master summaries, the investigator devised a content scoring rubric (Appendix D) for each video. Each content scoring rubric had five main points discussed or mentioned in the video. Participant notes were scored on how accurately the five main points were covered. Each of the five main points was scored on a three-point scale (see Figure 1).

0	1	2
Unsatisfactory	Satisfactory	Complete

Figure 1. Three point scoring scale

A score of 0 indicates that the point was not covered in the summary, a score of 1 indicated that the summary in part alluded to that detail and a score of 2 implied that that particular detail was covered in the summary. The content scoring rubric rated each summary on five main points. Each point could score a minimum of 0 and maximum of 2. After rating the entire summary on five main points, the scores of five main points were added to arrive at a final score. The final score for a summary could range from a minimum of 0 and a maximum of 10 points.

In this study the research question being investigated was which input modality would result in better notes. The dependent variables measured for each note were:

- Accuracy of content: To measure the accuracy of the content, a content scoring rubric (Appendix E) was employed to assess the summaries.

11

- Mental Load: The NASA TLX index (Appendix F) was employed to gather participants' perceptions about stress in using each mode. The task load index survey was administered after each trial.
- Preferred Mode: Post session, the questionnaire inquired about the preferred mode of input and why that mode was preferred (Appendix H).
- Number of words: The number of words was obtained by counting the words used to make up a summary.

Participants

Forty-three adults, native speakers of American English, between the ages of 25 - 55 participated in the study. Minimum education level of each participant was at least a college degree and all of them had day jobs. All participants used smartphones to make phone calls, text, and view or send emails. They were also familiar with other applications on the phone like voice memo, notes, and calendar application. The population gender split was 24 females and 19 males. Since all participants were at least college educated, the differences in their ability to describe the content of the video were deemed insignificant, as we assumed that all had similar comprehension and communication abilities and the tasks they performed were no more complex than they normally perform. The results of two participants were excluded from the final analyses, as the voice recording was not audible. The final analyses were based on the data collected from 41 participants. The number of participants for the experiment was based on calculations using the G*Power software based on a repeated measures experimental design, with a moderate power size of 0.81.

Apparatus and Materials

The devices used in this study consisted of an iPhone 4s, model MC922LL/A; Dell OptiPlex 755 PC, model EN-W7P64-7.2.00.0 with Intel Core 2 processor/64 bit system; Dell Keyboard RT7D50, 104 key English Keyboard; Dell Optical Mouse model M0C5U, USB Scroll 3; Dell LCD Flat Panel Monitor Model 1907FP, 19-inch screen size, 1280 x 1024 resolution. Additional materials consisted of paper and pen for handwritten summaries, paper copies of briefing scripts, participant agreements, post-trial NASA TLX surveys and post-study preference questionnaires.

For each trial, the participant launched the video by clicking on the link on the desktop. The participant used a mouse to resize the display window and to control the volume. After watching the video and depending on the modality selected, the participant used the "Voice Memos" application on the iPhone 4s for recording spoken notes, the "Notes" application on the iPhone 4s for typing notes, or paper and pen for writing notes by hand.

Procedure

Each participant was tested individually at the Oracle usability labs. Before each experiment the investigator prepared all equipment and materials needed. In addition, the environment was adjusted as needed to make sure the participants performed the tasks in the same testing conditions. After arrival, the investigator escorted the participant to the lab. In the lab, the participant read, agreed with and signed the consent form (Appendix B). Each participant was verbally briefed about experiment procedure and the tasks he or she would perform during the session. The investigator read the instructions from a script

(Appendix I) to make sure every participant received the same instructions. Then the participant randomly selected a video to watch and mode to summarize (see Figure 2 & 3).
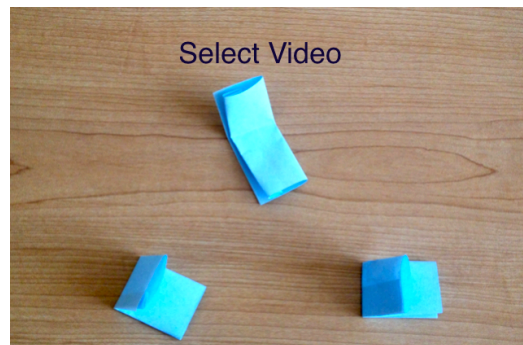
Figure 2. Randomly select a video
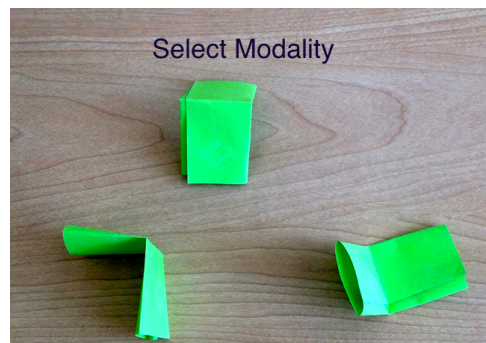
Figure 3. Randomly select a modality

The order of the videos and the modality used to record notes were randomized to reduce order bias. The participant then clicked the link on the desktop to launch the video on a PC. The participant watched the video once and was not allowed to take notes while viewing (see Figure 4). After watching the video, the participant was given two to three minutes to gather their thoughts. The investigator instructed each participant to record the

summary to the best of their ability, turned on a timer for three minutes and left the

participant alone to complete the summary. The investigator waited in the control room

until the participant completed the summary. The participant signaled the task completion

by raising an arm. The investigator returned to the experiment lab to administer the

NASA-TLX survey. Each participant received a five-minute break between experimental

runs. The sequence was repeated three times. After three trials, the participant completed

the preferred mode questionnaire. Finally, the investigator responded to any questions the

participants had about the experiment. The table below (Table 1) lists all steps of the

protocol and time allocated for each step.

Figure 4. Participant watched video here

Table 1

Steps of the protocol and time allocated for each step

Steps	*Tasks and time allocation*
Enrollment, Set Up and Calibration	• Introduction, consent form, debrief about experiment overview • Total time allotted = 15 minutes

No break

Tasks	• Randomly select a video and input modality. • Watch video once. Participants were not allowed to take notes while watching the video. • Allocate two minutes for participants to gather their thoughts. • Set timer to three minutes for participants to summarize the video. • Investigator leaves the participant alone to summarize. • After the participant is done summarizing, administer workload questionnaire. • Five-minute break between trials. • Repeat above protocol for two additional trials. • Total time allocated = 45 minutes.

No break

Post Session	• Preference questionnaire • Total time allotted = 5 minutes

No break

Debriefing	• Debriefing statement provided

Each participant produced three summaries – spoken, typed and handwritten. In all, 43 spoken, 43 typed, and 43 handwritten summaries were collected during the

experiment. The investigator transcribed 129 summaries in all in an Excel document after the experiment. The investigator compiled all summaries to be graded in a separate Excel document. The audio recordings of the two participants were not of good quality hence their data was excluded and the remaining123 summaries were included in the final analysis. Four graders, two males and two females, all fluent speakers of American English, were selected to rate the summaries. All four graders were at least college graduates. Multiple graders were used to reduce inter-rater bias. The grading document was mailed to them along with grading instructions (Appendix G). Each grader rated 123 summaries and mailed the documents back to the investigator. The final score for each summary was devised by taking an average of all four grader ratings. The investigator transferred the survey data into the Excel spreadsheet for further analysis.

Results

Two separate one-way repeated-measures analyses of variance (RM ANOVAs) were conducted to assess the effect of modality on accuracy of content and number of words in spoken, typed and handwritten levels.

The data were analyzed by first checking inter-rater reliability using the Pearson product-moment correlation coefficient. Next, Mauchly's test of sphericity was used to check equality of variances. RM ANOVA was used to do trend analysis based on preferred mode. And finally, the survey data of perceived mental load collected on an ordinal Likert-like scale was analyzed by performing RM ANOVA on all the sub-scales. While there are reservations in some fields of sciences regarding analyzing ordinal data using inferential statistics (Knapp, 1990; Jamieson, 2004), the use of RM ANOVA to analyze task load data load is common (Geoff, 2010).

Grading Reliability

Four graders rated the summaries using the content-scoring rubric prepared by the investigator. The four graders were not aware of the purpose of the experiment; they simply received the summaries, rating rubric, and grading instructions from the investigator. A Pearson correlation coefficient was computed for grader ratings to appraise the inter-grader reliability. A moderate to strong positive correlation (Table 2) among the ratings of all four graders indicated that the summary ratings were consistent in the same direction across all graders.

Table 2

Pearson correlation for inter-rater reliability (n = 41)

Grader	Statistic	Grader 2	Grader 3	Grader 4
Spoken				
Grader 1	Pearson Correlation	.520	.480	.594
	Sig. (2-tailed)	*<.001*	*<.01*	*<.001*
Grader 2	Pearson Correlation		.623	.456
	Sig. (2-tailed)		*<.001*	*<.01*
Grader 3	Pearson Correlation			.760
	Sig. (2-tailed)			*<.001*
Typed				
Grader 1	Pearson Correlation	.575	.495	.580
	Sig. (2-tailed)	*<.001*	*<.01*	*<.001*
Grader 2	Pearson Correlation		.472	.537
	Sig. (2-tailed)		*<.01*	*<.001*
Grader 3	Pearson Correlation			.533
	Sig. (2-tailed)			*<.001*
Handwritten				
Grader 1	Pearson Correlation	*.607*	*.728*	*.738*
	Sig. (2-tailed)	*<.001*	*<.001*	*<.001*
Grader 2	Pearson Correlation		*.658*	*.560*
	Sig. (2-tailed)		*<.001*	*<.001*
Grader 3	Pearson Correlation			*.716*
	Sig. (2-tailed)			*<.001*

Accuracy of Content

Mauchly's test of sphericity for the dependent variable quality indicated that the assumption of sphericity has been satisfied. A one-way RM ANOVA was conducted to

compare the effect of modality on the accuracy of content and there was a statistically

significant effect of modality on the accuracy of content, $F(1,40) = 10.65$, $p = .002$.
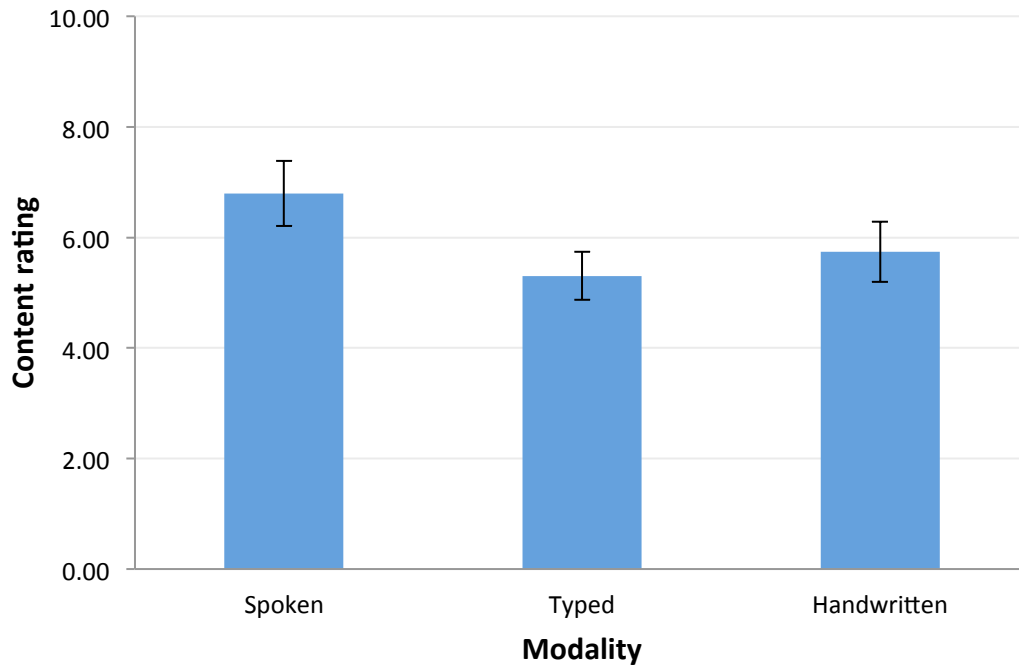

Figure 5. Average rating for accuracy of content

Glancing at the descriptive statistics reveals that the quality scores of spoken

notes were better than typed or handwritten notes (see Figure 5). A post-hoc Sidak test

revealed that the difference in the accuracy of spoken ($M = 6.80$, $SD = 1.87$) and typed

($M = 5.30$, $SD = 1.38$) notes was statistically significant at $p < .001$; the difference

between the accuracy of spoken ($M = 6.80$, $SD = 1.87$) and handwritten ($M = 5.70$, $SD =$

1.73) notes was also statistically significant at $p = .007$. Spoken notes were more accurate

compared to typed and handwritten notes. Another post-hoc Sidak test comparing

accuracy of typed ($M = 5.30$, $SD = 1.38$) and handwritten ($M = 5.7$, $SD = 1.73$) notes was

found to be statistically non-significant at $p = .55$. The accuracy of handwritten notes was no better than typed notes; alternatively, the accuracy of handwritten notes was as good as or comparable to the accuracy of typed summaries.

Mental Task Load

The NASA TLX required participants to report their experience level of mental load on six sub-scales: mental demand, physical demand, temporal demand, effort, performance and frustration. The self-reported mental workload was measured on a seven-point scale [Very Low (1) – Very High (7)] for all three modalities. A one-way RM ANOVA indicated that modality had a statistically significant effect on perceived mental load, $F(1, 40) = 867.86$, $p = .000$. This suggests that modality used in the experiment significantly impacted perceived task load.

A post-hoc Sidak test of the sub scales revealed that the physical demand in spoken condition ($M = 2.29$, $SD = 1.57$) was significantly lower than the physical demand in typed condition ($M = 4.00$, $SD = 2.09$, $p = .000$). Furthermore, the physical demand in spoken condition ($M = 2.29$, $SD = 1.57$) was significantly lower than the handwritten condition ($M = 3.78$, $SD = 1.85$, $p = .000$).

A post-hoc Sidak pairwise comparison of temporal demand, another sub scale, indicates that temporal demand placed during spoken condition ($M = 2.73$, $SD = 1.67$) was significantly lower than that placed during typed condition ($M = 4.41$, $SD = 1.62$, $p = .000$) as well as handwritten condition ($M = 4.34$, $SD = 1.49$, $p = .000$) (see Figure 6).
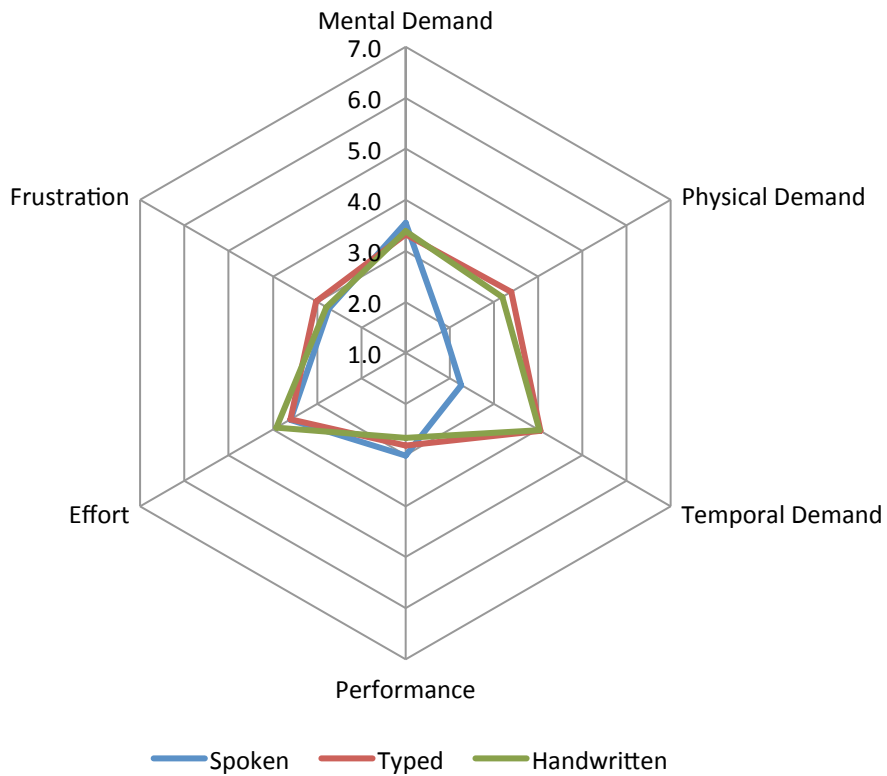
Figure 6. Mental task load measured using NASA TLX

Preferred Mode Trend Analysis

At the end of the session, after interacting with all three modalities, participants were asked to vote for their preferred mode of taking notes, and qualitative data around the same. Analysis of preference data revealed that 44% of the participants preferred spoken mode to take notes, 32% preferred typed notes and 24% preferred handwritten notes. Tables 3, 4, and 5 list some of the reasons participants preferred a particular mode.

Table 3

Participant comments for preferring spoken modality

Reasons for preferring spoken mode
"Easy to use. Didn't worry about spelling and legibility of writing."
"Easy method, freedom to express more easily and add to description as I went along."
"I feel less pressured…speech allows me to get everything out accurately and quickly."
"Speaking requires less effort for me."
"Speaking is faster…my handwriting is messy…typing is annoying."

Table 4

Participant quotes for preferring typed modality

Reasons for preferring type mode
"Typing allows me to structure and edit what I have to say."
"I type fast…it is natural for me. I can visually look over my notes at a later date."
"I seem to think and output at the same time…speaking seemed like I was rambling."

Table 5

Participant quotes for preferring handwritten modality

Reasons for preferring handwritten mode
"Writing by hand allowed me to express my thoughts immediately without obstacles."
"...While typing I spent more time correcting errors versus getting summary out."
"I am more comfortable with writing to describe my thoughts than I am speaking extemporaneously."

Examining the comments shows that participants had strong inclinations for one mode over the others. Partitioning the data by preference and looking at descriptive statistics shows that spoken notes were more accurate and detailed regardless of preference (see Figure 7 & Figure 8); however, preference did impact the accuracy of content and number of words captured for typed and handwritten notes. Those who preferred typed mode wrote more accurate and detailed notes while typing, whereas those who preferred handwritten mode produced accurate and detailed notes when writing by hand. The differences between the two preferred modes were statistically non-significant.
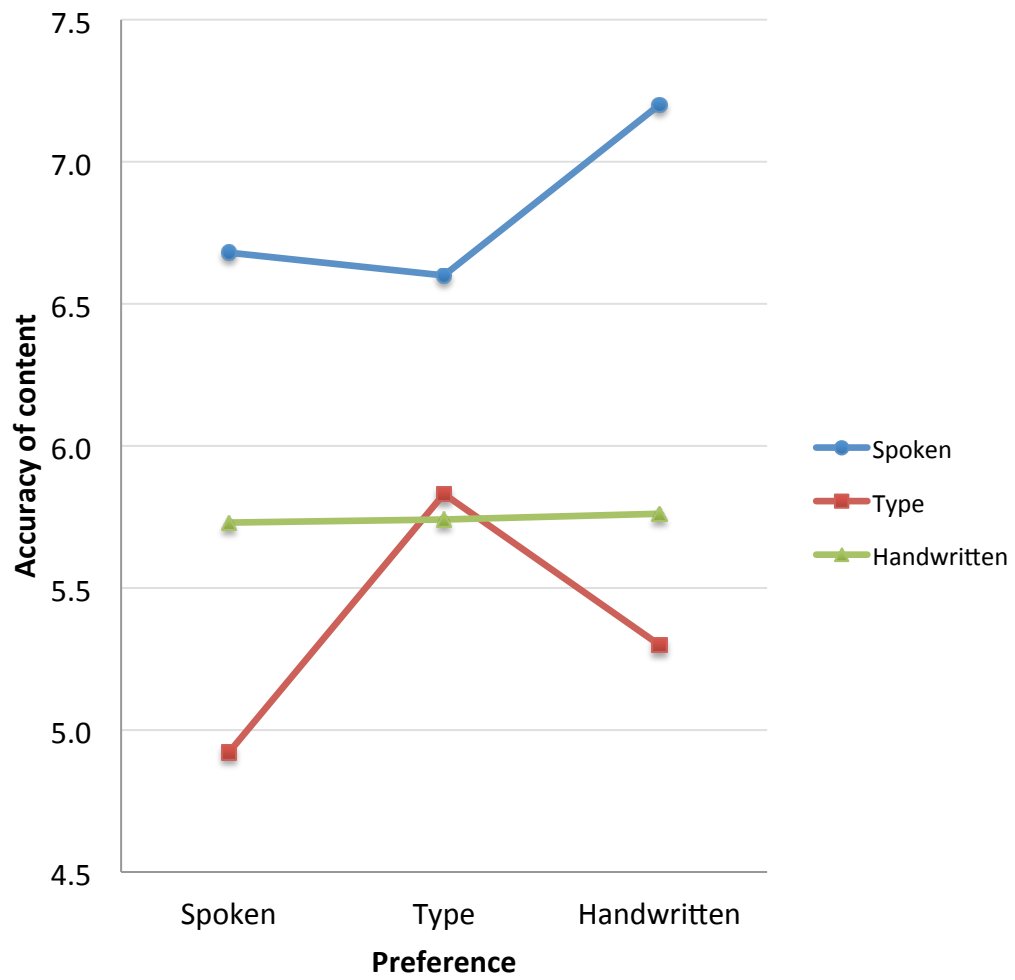
Figure 7. Accuracy of content grouped by preference
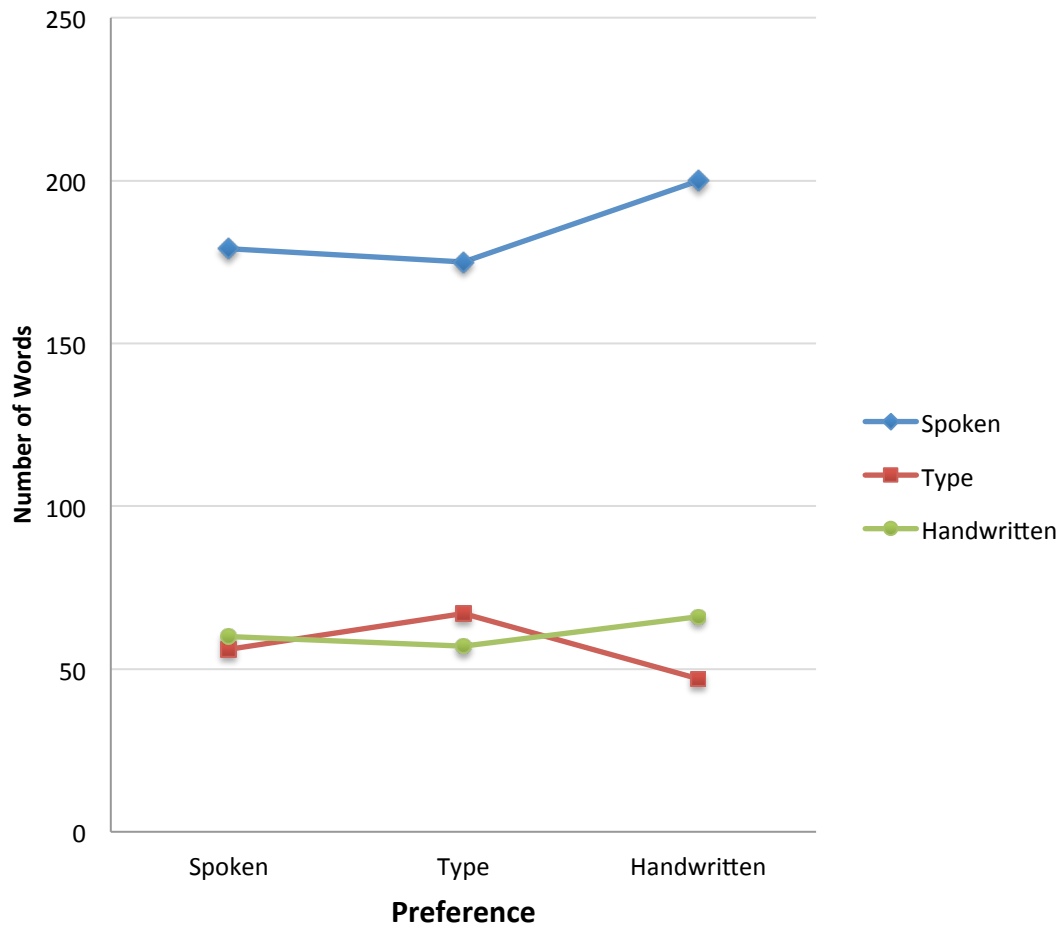
Figure 8. Number of words grouped by preference

Number of Words

Mauchly's test of sphericity for the dependent variable number of words indicated that the assumption of sphericity has not been met, $\chi^2(2) = 61.05$, $p = 0.000$. The degrees of freedom were corrected using the Greenhouse-Geisser estimates of sphericity ($\varepsilon = .56$). There was a statistically significant effect of modality on the number of words, $F(1.12,$

44.65) = 114.51, p = .000. The findings suggest that the type of modality has significant effect on the amount of details produced.
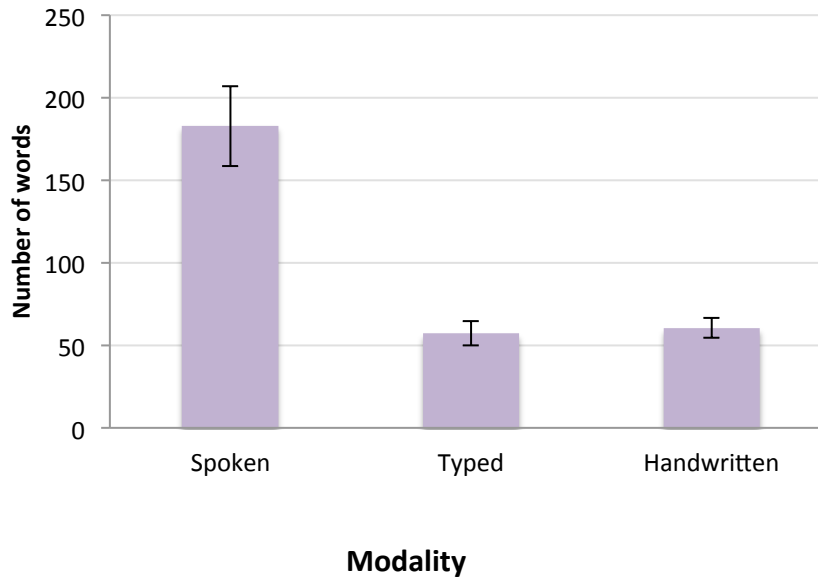

Figure 9. Average number of words by modality

Inspection of the descriptive statistics for number of words reveals that spoken notes were more detailed than typed or handwritten (see Figure 9). A post-hoc Sidak test illustrates that the difference between the number of words for spoken notes (M = 182.83, SD = 76.83) and typed notes (M = 57.37, SD = 23.47) was statistically significant at p = .000; the difference between the number of words captured for spoken notes (M = 182.83, SD = 76.83) and handwritten notes (M = 60.56, SD = 19.04) was also statistically significant at p = .000. The post-hoc Sidak test for the difference in number of words captured for typed (M = 57.37, SD = 23.47) and handwritten (M = 60.56, SD = 19.04)

notes was statistically non-significant at $p = 0.98$.   The results suggest that spoken notes were more detailed and the typed and handwritten summaries were more economical. Additionally, the amount of details produced in typed and handwritten mode was about the same.

Discussion

The goal of this study was to investigate the effects of modality on notes produced. Three modalities were evaluated: spoken, typed, and handwritten. The dependent measures were accuracy of content, mental task load, user preference, and number of words. Based on the review of literature it was anticipated that spoken notes would likely result in better accuracy and incur lower task load, and this in turn would inspire participants to provide more detailed summaries as compared to typed or handwritten notes.

The study findings supported these hypotheses. Spoken notes were not only more detailed, they also captured the gist of videos more accurately as compared to typed or handwritten notes. A major distinction between spoken modality versus typed/handwritten modality was the dimension of progress; sound is perceived in time dimension whereas typed and handwritten modes are acknowledged spatially (McClelland & Elman, 1986). The process of typing on mobile devices or writing on paper by hand requires sentences to be composed mentally by organizing words in grammatically correct sequences, binding of character shape to the alphabetical character and complex motor programs to be initiated to make specific hand movements to complete the task (Flower & Hayes, 1981). In short, the elaborate process of typing or writing by hand places substantial demands on working memory (Kellogg, 1996). Speaking, meanwhile, not only bypasses spelling and character shape association to free up space in working memory, it also has additional resources like Broca's and Wernicke's area to aid speech production.

Objectively, looking at the raw transcribed data, many segment fragments and incomplete phrases were noticed in notes captured using spoken mode. Overall, more participants preferred typed or handwritten notes (56%) as compared to spoken notes (44%). A majority of participants preferred typed or handwritten mode as they could see what had been written so far, whereas in the spoken mode there was no such affordance, as in, "I feel like repeating myself, as I don't remember what I have said so far." Another reason handwritten and typed modes were preferred was the ability to reiterate notes and revise and edit.

It was expected that typed mode would likely result in more accurate notes and be more detailed as compared to handwritten mode. Those hypotheses were rejected. Typed mode was presumed to produce more accurate and detailed notes than handwritten, as one-step recognition is used in typed mode and two-step recall is used in handwritten mode. One potential reason is that even though recognition is faster, tapping on a small keyboard takes longer (Fitts, 1954). Faster typing speed didn't matter either as participants took time to correct errors, edit and reiterate.

In the case of a preferred mode, spoken notes were by far better in terms of accuracy and amount of details regardless of preferred mode. However, preference seemed to have an impact on typed and handwritten notes. Those who preferred to type found writing by hand cumbersome and physically painful, whereas those who preferred writing by hand found it inconvenient to type on a small keyboard and deemed autocorrect annoying. Apart from the complex writing process, choosing the right tool to transcribe written content can impact the output. Future studies must look into evaluating

handwriting software as well.

Conclusion

The premise of the study was to investigate modalities and find the one best suited for in-field sales representatives. Even though spoken notes were better and had more details than typed or handwritten notes, many of them had sentence fragments and incomplete thoughts similar to rough notes. Spoken notes are perhaps best for a certain context like quick capture of data meant to be referenced by the author in the future or for jotting down important points before they escape memory.

Spoken notes make sense in a sales-rep scenario where the rep might want to capture important meeting points before heading to another meeting. It also makes sense in the case of an eyewitness scenario where a subject is expected to describe or recall a witnessed event and detail is more important than a complete sentence. In such an example, spoken recall might result in less loss of details as compared to typing or writing by hand. Spoken notes can also be useful during doctor-patient meetings. Recording spoken notes of patients might be a more efficient modality for capturing maximum details. Spoken mode can also be useful in cases of subjects with writing difficulties.

Another point of interest is the audience for the notes. In spite of the fact that spoken notes were far better than typed or handwritten notes, a majority of the participants (56%) preferred to type or write notes by hand. Perhaps this was because they knew the data would be evaluated by someone, in this case the experimenter, and hence preferred to type or write by hand as they could produce polished material.

Limitations and Future Research

A major limitation of the study was the length of the video, three minutes. Meetings that happen in the work environment are generally longer. Moreover, a three-minute time limit was established to produce spoken, typed or handwritten notes. In the lab experiment multiple constraints were imposed and variables were measured in an isolated environment to prevent confounds. Taking notes using spoken modality in the real environmental would introduce background noise. This could be a concern when using speech modality in the natural environment and must be addressed in future studies. Privacy could be another aspect that is compromised when using spoken mode in a natural environment. Another major limitation of the study was comparing handwritten notes to typed notes. Currently, many handwriting recognition software applications are available on the market that allows a user to scribble notes directly on the screen. Future studies must consider evaluating that as one of the modalities. This study did not measure grammatical or spelling errors as an auto correct setting was used for typing on the mobile device. Future studies might figure out ways to take that into account as well. Several aspects need further investigation, such as the possibility that preference might play a role in the accuracy of notes and how detailed the notes are. The type of audience can also impact the quality of notes. Future studies must be conducted to address these concerns.

References

Baddeley, A. D., & Hitch, G. (1974). Working Memory. *The psychology of learning and motivation: Advances in research and theory , 8*, 47-89.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science , 4* (11), 417-423.

Barnard, L., Yi, J., Jacko, J. A., & Sears, A. (2005). An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices. *International Journal of Human-Computer Studies , 62* (4), 487-520.

Basapur, S., Xu, S., Ahlenius, M., & Lee, Y. S. (2007). User expectations from dictation on mobile devices. *In Human-Computer Interaction. Interaction Platforms and Techniques* (pp. 217-225). Springer Berlin Hiedelberg.

Benjamin, A. (2009, Feb). *Arthur Benjamin: Teach statistics before calculus!* Retrieved June 01, 2014, from Ted Talks: http://www.ted.com/talks/arthur_benjamin_s_formula_for_changing_math_education?language=en

Bi, X., Li, Y., & Zhai, S. (2013). FFitts law: modeling finger touch with fitts' law. *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1363-1372). ACM.

Bijani, C., White, B.-K., & Vilrokx, M. (2013). Giving voice to enterprise mobile applications. *In Proceedings of the 15th international conference on Human-*

computer interaction with mobile devices and services (pp. 428-433). Munich: ACM.

Bourdin, B., & Fayol, M. (2000). Is graphic activity cognitively costly? A developmental approach. *Reading and Writing , 13* (3-4), 183-196.

Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and brain sciences , 24* (01), 154-176.

Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and brain sciences , 22* (05), 735-808.

Fitts, P. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology , 47* (6), 381.

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College composition and communication ,* 365-387.

Geoff, N. (2010). Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education , 15* (5), 625-632.

Gould, J., Greene, S. L., Boies, S. J., Meluson, A., & Rasamny, M. (1990). Using a touchscreen for simple tasks. *Interacting with computers , 2* (1), 59-74.

Hoggan, E., Brewster, S. A., & Johnston, J. (2008). Investigating the effectiveness of tactile feedback for mobile touch screens. *In Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1573-1582). ACM.

Jacquemot, C., & Scott, S. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in cognitive sciences , 10* (11), 480-486.

Jamieson, S. (2004). Likert Scales: how to (ab) use them. *Medical Education , 38* (12), 1217-1218.

Kellogg, R. T. (1996). A model of working memory in writing. 55-71.

Kellogg, R. T., Olive, T., & Piolat, A. (2007). Verbal, visual, and spatial working memory in written language production. *Acta Psychologica , 124* (3), 382-397.

Kintsch, W. (1970). Models for free recall and recognition. *Models of human memory* , pp. 331-373.

Knapp, T. R. (1990). Treating ordinal scales as interval scales: an attempt to resolve the controversy. *Nursing research , 39* (2), 121-123.

Lieberman, p. (1993). *Uniquely human: The evolution of speech, thought, and selfless behavior.* Boston: Harvard University Press.

Lyons, K., Plaisted, D., & Starner, T. (2004). Expert chording text entry on the twiddler one-handed keyboard. *In Wearable Computers, 2004. ISWC 2004. Eighth International Symposium. 1*, pp. 94-101. IEEE.

Maxwell, J. P., Masters, R. S., Kerr, E., & Weedon, E. (2001). The implicit benefit of learning without errors. *The Quarterly Journal of Experimental Psychology: Section A , 54* (4), 1049-1068.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology , 18* (1), 1-86.

Moore, T. (2012, Feb). *Why is 'x' the unknown?* Retrieved June 01, 2014, from www.ted.com: https://www.ted.com/talks/terry_moore_why_is_x_the_unknown/transcript?language=en

Nass, C., & Gong, L. (2000). Speech interfaces from an evolutionary perspective. *Communication of the ACM , 43* (9), 36-43.

Newell, A., & Card, S. K. (1985). The prospects for psychological science in human-computer interaction. *Human-computer interaction , 1* (3), 209-242.

Parhi, P., Karlson, A., & Bederson, B. (2006). Target size study for one-handed thumb use on small touchscreen devices. *In Proceedings of the 8th conference on Human-computer interaction with mobile devices and services* (pp. 203-210). ACM.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory Illusions: Recalling, recognizing, and recollecting events that never occured. *Journal of Memory and Language , 35* (2), 261-285.

Price, K., Lin, M., Feng, J., Goldman, R., Sears, A., & Jacko, J. (2006). Motion does matter: an examination of speech-based text entry on the move. *Universal Access in the Information Society*, *4* (3), 246-257.

Rampoldi-Hnilo, L., White, B.-K., Snyder, M., & Sampanes, C. (2009). The future of enterprise is with the mobile workforce: An International Field Study. *In Internationalization, Design and Global Development* (pp. 308-315). Springer Berlin Heidelberg.

Rivera, J., & Van Der Meulen, R. (2014, February 13). Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013. Egham, UK.

Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological bulletin*, *95* (3), 355.

Sears, A., Revis, D., Swatski, J., Crittenden, R., & Shneiderman, B. (1993). Investigating touchscreen typing: the effect of keyboard size on typing speed. *Behavior & Information Technology*, *12* (1), 17-22.

St. John, R. (2005, Feb). *Richard St. John: 8 secrets of success.* Retrieved June 01, 2014, from www.ted.com:
http://www.ted.com/talks/richard_st_john_s_8_secrets_of_success?language=en

Tamminen, S., Oulasvirta, A., ToisKallio, K., & Kankainen, A. (2004). Understanding mobile contexts. *Personal and ubiquitous computing*, *8* (2), 135-143.

Wickens, C. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society , 50* (3), 449-455.

Wickens, C. (2002). Multiple resources and performance prediction. *Theoritical issues in ergonomics science , 3* (2), 159-177.

Winstein, C., & Schmidt, R. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Pyschology: Learning, Memory, and Cognition , 16* (4), 677-691.

Appendix A

Human Subjects – IRB Approval

To: Dr. Sean Laraway
Department of Psychology
San Jose State University
One Washington Square
San Jose, CA 95192-0120

Chaya Bijani

From: Pamela Stacks, Ph.D.
Associate Vice President
Graduate Studies and Research

Date: May 20, 2014

The Human Subjects-Institutional Review Board has approved your
request to use human subjects in the study entitled:

"Effect of input modality on captured data"

This approval is contingent upon the subjects participating in your
research project being appropriately protected from risk. This includes the
protection of the confidentiality of the subjects' identity when they
participate in your research project, and with regard to all data that may be
collected from the subjects. The approval includes continued monitoring
of your research by the Board to assure that the subjects are being
adequately and properly protected from such risks. If at any time a
subject becomes injured or complains of injury, you must notify Dr.
Pamela Stacks, Ph.D. immediately. Injury includes but is not limited to
bodily harm, psychological trauma, and release of potentially damaging
personal information. This approval for the human subject's portion of
your project is in effect for one year, and data collection beyond
May 20, 2015 requires an extension request.

Please also be advised that all subjects need to be fully informed and
aware that their participation in your research project is voluntary, and that
he or she may withdraw from the project at any time. Further, a subject's
participation, refusal to participate, or withdrawal will not affect any
services that the subject is receiving or will receive at the institution in
which the research is being conducted.

If you have any questions, please contact me at (408) 924-2427.

Protocol #S1402076

Appendix B

Consent Form

Agreement to Participate in Research

Responsible Investigator(s): Chaya Bijani, graduate student, San Jose State University
Title of Protocol: Effect of modality on captured data.

1. You have been asked to participate in an empirical study on the effect of input modality while recording summarized details.

2. You will be asked to watch a video and summarize the details of the video using speech, type or handwritten mode. You will be asked to fill a workload questionnaire after every trial run. The study will last approximately 2 hours and will be conducted in Oracle Usability Labs.

3. Participation in this study involves minimal to no risk. The probability and magnitude of harm or discomfort are no greater than would be encountered in daily life. Although the results of this study may be published, no information that could identify you will be included.

4. You will not receive any direct benefits for participating in this study. You will be compensated for your time with iTunes gift card of $5- $10. No other compensation is provided for participation in the study.

5. Questions about this research may be addressed to Chaya Bijani at chaya.bijani@sjsu.edu. Complaints about the research may be presented to Dr. Sean Laraway, Associate professor, Graduate Program in Human Factors/Ergonomics, Department of Psychology at (408) 924-5679. Questions about a research subject's rights, or research-related injury may be presented to Pamela Stacks, Ph.D., Associate Vice President, Graduate Studies and Research at (408) 924-2427.

6. No service of any kind, to which you are otherwise entitled, will be lost or jeopardized if you choose not to participate in the study. Your consent is being given voluntarily. You may refuse to participate in the entire study or in any part of the study. You have the right not to answer questions you do not wish to answer. If you decide to participate in the study, you are free to withdraw at any time without any negative effect on your relations to San Jose State University. At the time you sign this consent form, you will receive a copy of it for your records, signed and dated by the investigator.

The signature of the participant on this document indicates agreement to participate in the study. The signature of the researcher on this document indicates agreement to include the above named subjects in the research and attestation that the subject has been fully informed of his or her rights.

_____   _____
Signature of Participant   Date

_____   _____
Signature of Investigator   Date

Participant Recruitment Flyer

Participate in Research!

What

We are looking for volunteers to participate in a study about various input modalities for phone device. The study will take approximately 30 min to complete.

Who

Age group 25 – 55yrs
Other Requirement - Native speaker of American English

Where

Oracle Usability Labs (Bldg 1, 2nd floor),
5805 Owen Drive, Pleasanton

Compensation

In appreciation of your time, you will receive an *iTunes Gift Card($10) or an Oracle branded gift*

If you are interested, please contact
Chaya Bijani
Chaya.bijani@oracle.com

Thank you!

Appendix D

Content-Scoring Rating Rubric

The Unknown X

Topic	Detail 1	Detail 2	Detail 3	Detail 4
Why "X" used to represent unknown entities in Western Culture?	The speaker details how math, logic and engineering flourished in Arabic culture during the 11th and 12th centuries.	The Spanish were interested in this wisdom, but there were issues translating the Arabic term "Shay" used to represent the "unknown" in math proofs.	The "Shay/Sh" sound does not exist in Spanish, so the Spanish replaced the "Sh" sound by the Greek symbol chi/Kai.	When Greek literature was translated into Latin, the "chi" symbol becomes the letter "X." Latin textbooks were used for almost 600 years in Western culture.

Steps to Success

Topic	Detail 1	Detail 2	Detail 3	Detail 4
What leads to success? Or how to be successful? Or how to achieve success?	A young student asked him what leads to success, and the speaker interviewed 500 successful people over 7 years to find the answer	The speaker found 8 basic principles that lead to success	Any 4 of the following 8 principles: Passion, Work, Focus, Good/Practice, Push, Serve, Ideas, Persist	Rest 4 of the 8 principles: Passion, Work, Focus, Good/Practice, Push, Serve, Ideas, Persist

Statistics vs. Calculus

Topic	Detail 1	Detail 2	Detail 3	Detail 4
The topic is about math curriculum Or it proposes to change math education system in the US. Or the topic is about Statistics versus calculus.	Learning Statistics is more important than learning Calculus. The summit of the pyramid must change from calculus to statistics	Calculus is important but not useful in everyday life.	Statistics is relevant in everyday life. It is practical, fun, and engaging at the same time as it teaches us about risks, rewards, games, and gambling.	The world has changed from analog to digital. It's time for our math curriculum to change from classic, continuous math to modern, discrete math.

Appendix E

NASA TLX Workload Index

Date#_____ Run#_____
Participant #_____ Condition_____

SARDA Workload Questionnaire

Workload: Please evaluate the workload of your tasks in the simulation by circling the value that represents the level of workload you felt OR drawing a line at the point on the scale that matches your experience.

1. **Mental Demand**: How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

    Low   1   2   3   4   5   6   7   **High**

2. **Physical Demand**: How much physical activity was required (e.g., radio communications, button presses, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous?

    Low   1   2   3   4   5   6   7   **High**

3. **Temporal Demand**: How much time pressure did you feel due to the rate or pace at which the task or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

    Low   1   2   3   4   5   6   7   **High**

4. **Performance**: How successful do you think you are in accomplishing the goals of the task? How satisfied are you with your performance in accomplishing these goals?

    Not Satisfied   1   2   3   4   5   6   7   **Very Satisfied**

5. **Effort**: How hard did you have to work (mentally and physically) to accomplish this level of performance?

    Didn't Have To Work   1   2   3   4   5   6   7   **Worked Very Hard**

6. **Frustration**: How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent do you feel in performing the task?

    Very Relaxed   1   2   3   4   5   6   7   **Very Discouraged**

Appendix F

Grading Instructions

Rules to grade the summary.

The purpose of grading is to rate the quality of the summary. A good summary is expected to address five elements: a topic sentence and four details in support of the topic sentence.

The 3-point rating scale is ("0" = Unsatisfactory; "1" = Satisfactory; "2" = Complete) is used to rate the quality of summary.

In the Excel sheet, the summaries are listed on the vertical, and the five elements are displayed across on the horizontal. I want the grader to evaluate how far/close the summary is from the five elements.

Each element can score 0, 1 or 2 points. The rating scale usage:

> 0=Unsatisfactory, means the element/detail in question is not mentioned in the summary.
>
> 1=Satisfactory, means the summary alludes/refers/makes a partial mention to the element/details in question.
>
> 2=Complete, means the element/detail can be clearly inferred from the summary.

A summary can get a maximum score of 10. Maximum two points can be allocated per element addressed.

Appendix G

Post-Study Preference Questionnaire

User Preference

1. Modality Preference: Which mode of input was most preferred?

    a. Spoken, Typed, or Handwritten

2. Why