

# Integrating Speech and Touch to Alleviate Screen Size and Input Constraints in Mobile Devices

*Jesse Cirimele, Scott Klemmer*

Stanford University Computer Science

353 Serra Mall

Stanford, CA, 94305, USA

[Cirimele@stanford.edu](mailto:Cirimele@stanford.edu), [SRK@stanford.edu](mailto:SRK@stanford.edu)

## ABSTRACT

In this project we surveyed people about their mobile email use and built paper prototypes to explore integrating touch and speech within mobile email. We found that users were interested in our design alternatives that had multi-modal speech and touch components for email composition and error recovery as well as for cross application tasks such as adding a calendar event from within the email application. Our future work is to build an interactive mobile version of email and calendar to further explore this task space.

## INTRODUCTION

As computers become more compact and more mobile, and as mobile phones become more powerful, the techniques that we use to interact with them must also evolve. With all these devices designers must deal with limited screen space, and most current devices text input is a virtual or physical thumb keyboard. One possible technique to alleviate these limitations is through multimodal speech and touch.

While people can talk much faster than they type [2], speech recognition performance falls short of desktop keyboard entry when time for error correction is factored in [7][3]. Yet these speeds would improve on current mobile entry, where the use of multimodal (speech and keypad) input can alleviate some of this strain by leveraging the strengths of each of these input styles [2][6].

In order to become familiar with our application area, email, we started off with surveys to see how people currently use email on their mobile devices.

## SURVEYS

We had participants carry around note cards for four days. They would mark down each time they checked their email, how many they received, read, and sent in each day. After the survey we informally asked users to reflect on their usage patterns over the period. We wanted this information to establish a baseline of expected use for mobile email.

## Results

Copyright is held by the owner/author(s).  
UIST'09, October 4–7, 2009, Victoria, BC, Canada.

Users checked their email around 5 times per day and received anywhere from 0 to 58 emails on their mobile device. Of these users would on average read just under half of them on their mobile device. Users sent on average only a single email from their mobile device per day.

In our interviews we were interested in understanding why people didn't read all of their emails, and why they didn't send very many emails. Users often would cite difficulty with input and reading in comparison with their desktop as the main reason they didn't use mobile email more. Often when they sent messages it was only the message was too important to wait until they could use a desktop.

The interviews motivated us to look deeper into alternate mobile input techniques and at ways that mobile email might be used for a different purpose than desktop email. We decided to build a paper prototype.

## PAPER PROTOTYPE

We built paper prototypes to explore two different user tasks with one or more variations. Each user went through each of these tasks with each of the variations. Each user was given a sheet of possible voice and menu commands that they could use to navigate the prototype. Users were asked to use a combination of voice and touch tasks.

In the first task, users were asked to find an event in one of the emails in their inbox and create a calendar appointment from that event. There were three separate variations for this task. In one, a pop-up window appeared over the email, which allowed the user to enter title, date, and comments for the event. In a second, users added the task to the calendar with a freeform voice command. In the third, the email was shrunk to only take half of the screen and the other half of the screen had 5 days of a calendar (See Figure 1).

In the second task, users were asked to compose an email in response to any email that they found in the inbox of the mail program using their voice. We had users wear a headset that sent input to a desktop running Dragon Naturally Speaking software (DNS). We placed the paper email client over a USB touch screen tethered to that desktop computer so that we could display the recognized text.

## Results

We used our compose email task to get a baseline of what users expectations were for voice recognition technology.

We found most users were not aware of how advanced speech recognition is today, and were more excited at the possibility of voice enabled applications after running this task. Users were able to successfully compose their emails with few errors. This was helped by conditions being close to optimal for speech recognition, however it is encouraging to move forward and see how useful speech would be as an input method in the wild.

In the “adding an email event to the calendar” task users gave a clear preference for not overlapping pop-up windows over the email, but rather for the constraint-based approach that kept the email side-by-side with the window that came up. Users were very interested by the idea of popping up the full calendar week next to the email, as this allowed them to see the context of their schedule while they were choosing whether or not to add then event.

Since we weren’t using DNS for this part it was more difficult to critically compare the voice-enabled add to calendar to the non-voice enabled version on a technical level, but the idea of giving an add to calendar command with speech while looking at the email was easy for users to understand, use, and remember on subsequent variations.

In different variations, users were asked to use touch, speech, or both to navigate the interface. We found that even with a list of available commands, having important often used commands visible as buttons on screen was crucial. For less commonly used commands, however, users would often look at the sheet of possible commands and speak the voice command.

This indicated to us that for situations where you may be doing a task relatively often, but can’t afford the screen space to visualize all the commands used to complete the task, using speech to access these commands will prove effective. In particular, we feel tasks that cross application boundaries will fall into this category. In these tasks it is often that case that you can’t put all the commands from a second application into the first, but you could access them with speech and gain information from the context of the first application. If users also use these applications separately it will be very natural to connect them using the same voice commands across applications.

### CONCLUSIONS AND THE FUTURE

We found that multimodal interaction with touch and speech provides a viable alternative to switching between applications in cross-application tasks and as alternate input for email composition. We will continue to explore this, and mobile email organization and presentation with our interactive prototype as it develops.

In order to do this we would like to build a couple of versions of the add to calendar scenario in our interactive prototype so that we can explore factors that are abstracted out of the paper prototype such as time to complete tasks, recovering from errors, and contextual constraints.



Figure 1: A user comparing alternate versions in the “add event to calendar from email” task.

We believe that the “add to calendar” scenario is part of a broader space of interesting tasks that benefit from multi-modal interaction when performed on mobile devices. We will explore other cross-application tasks such as inserting an event into an email, adding contacts to an email or calendar event, and re-organizing the inbox to determine if this is the case.

### ACKNOWLEDGMENTS

We thank Nokia Research for their support.

### REFERENCES

1. M. Cohen, J. P. Giangiola, J. Balogh. “Voice User Interface Design.” Addison-Wesley, 2004.
2. A. Cox, P. Cairns, A. Walton, and S. Lee. Tlk or txt? using voice input for sms composition. *Personal and Ubiquitous Computing*.
3. J. Karat, D. Horn, C. Halverson, and C. Karat. Overcoming unusability: developing efficient strategies in speech recognition systems. *Conference on Human Factors in Computing Systems*, 2000.
4. S. Klemmer, A. K Sinha, J. Chen, J. Landay, N. Aboobaker, A. Wang, “SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces” *UIST 2000: ACM Symposium on User Interface Software and Technology*: pp. 1-10.
5. Lee, Kwan-Min and J. Lai. “Speech Versus Touch: A Comparative Study of the use of Speech and DTMF Keypad for Navigation”. *International Journal of Human Computer Interaction*.
6. S. Oviatt and P. Cohen. What comes naturally. *COMMUNICATIONS OF THE ACM*, 2000.
7. N. Patel, S. Agarwal, N. Rajput, A. Nanavati, P. Dave, T. Parikh. "A Comparative Study of Speech and Dialed Input Voice Interfaces in Rural India" In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2009)*.
8. B. Shneiderman. The limits of speech recognition: Understanding acoustic memory and appreciating prosody (2000). *drum.umd.edu*, 2005.