Errors? Not Too Worrisome. Exploring the Effects of Errors in Speech Transcription

by

Tianyi (Vera) Bao

B.A., University of British Columbia, 2021

Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Science

in the
School of Interactive Arts and Technology
Faculty of Communication, Art and Technology

Declaration of Committee

Name: **Tianyi (Vera) Bao**

Degree: **Master of Science**

Thesis title: **Errors? Not Too Worrisome. Exploring the Effects of Errors in Speech Transcription**

Committee: **Chair:** Jon Corbett
Instructor, Interactive Arts and Technology

Wolfgang Stuerzlinger
Supervisor
Professor, Interactive Arts and Technology

Brian Fisher
Committee Member
Professor, Interactive Arts and Technology

Marek Hatala
Examiner
Professor, Interactive Arts and Technology

Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University

or has conducted the research

c. as a co-investigator, collaborator, or research assistant in a research project approved in advance.

A copy of the approval letter has been filed with the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

Update Spring 2016

Abstract

This thesis presents two user studies that investigate how the errors that can occur during speech recognition affect users' text entry performance and their experience. For this work, I used a speech recognition system that injects believable errors in a controlled manner, and where users could fix errors simply by re-speaking a small part of their original utterance. Participants were asked to transcribe a set of phrases using the system, either with or without the insertion of errors, In the first study, I injected up to 33% errors, but saw no substantial differences. Yet, participants consistently commented on the used phrase set, which did not correspond well with spoken English. Thus, I created a novel phrase set based on spoken phrases. In the second study, I used this phrase set and inserted errors into the speech recognition results with either 25% or 50% probability. The results showed that inserting errors in the speech recognition system had a significant effect on participants' perceived mental workload. In addition, I found that inserting errors increased the number of errors users made during the task. More importantly, I discovered that the more erroneous the system was, participants were less likely to detect errors. The new phrase set for spoken English also performed substantially better than the one used in the first set. Finally, according to my findings, users have a fairly high tolerance towards errors encountered in speech transcription.

Keywords: human-computer interaction; speech recognition; text entry; error correction

Acknowledgements

I am immensely grateful to my supervisor Wolfgang for his invaluable guidance, expertise, and unwavering support throughout the entire research process. His constant encouragement has played a pivotal role in shaping this thesis. I am truly grateful for having him as my supervisor.

I am grateful for Dr. Marek Hatala for his thoughtful insights into my work that have contributed to the data analysis of this thesis. Special thanks to my former professor Dr. Benjamin Cheung, for supporting me both intellectually and emotionally.

I would like to extend my sincere gratitude to my family and friends for their support throughout my journey, especially friends from my 2021 SIAT cohort. They made the past two years so much more enjoyable. Another special thanks to *Kimetsu no Yaiba* for bringing me so much joy and mental strength.

I am also grateful to Afshan Ahmed for the initial prototype of the system, and Mohammad Rajabi Seraji for his invaluable help with fixing bugs in the system.

Lastly, I extend my appreciation to the members of my thesis committee.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In today's digital era, text entry has become an indispensable aspect of both my social and professional lives. Hence, efficient text entry has become vital for effective communication. Commonly used state-of-the-art technologies to improve text entry efficiency include auto-correction and word prediction. However, these features are not always perfect. In fact, some studies have shown that autocorrection and predictive features rarely increase text entry speed significantly, due to the time required to manually fix wrong predictions, corrections, and/or the higher cognitive load required to fix such errors [3, 5, 29, 38]. When such errors occur, users also experience an increase in frustration as well as physical and mental workload [3, 38, 39].

Beyond predictive features and autocorrection for typing-based input, voice-to-text has become another widely used modality of text entry [13, 45]. With recent advances in Automatic Speech Recognition (ASR) or Speech-to-Text (STT) technologies, voice assistant tools such as Google Home and Amazon Alexa have gained popularity, and today most digital devices support built-in dictation. A body of research has shown that entering text through voice can be much faster than typing [18, 26, 43, 50], albeit at the cost of more time-consuming error correction [30]. Yet, there has been a lack of studies that investigate how errors encountered during speech recognition affect users' text entry performance and also their frustration.

The central objective of my work is thus to understand how users respond to errors encountered in speech transcription tasks. Accordingly, I present a quantitative analysis of user responses to different error frequencies, complemented by qualitative user feedback and observations. In addition, based on feedback from users received when they encountered a traditional text entry phrase set, I also created a new phrase set that is (better) suited for speech transcription tasks.

Chapter 2

Related Work

In this section, I review prior work that is relevant for my work. I cover the basic mechanics of automatic speech recognition systems, error correction methods for voice-based text entry, the limits of human error detection, and metrics to measure text entry performance and efficiency.

2.1 Speech recognition technologies

Speech is one of the most natural means of communication between people [25]. The main goal of ASR systems is to make intelligent systems process spoken language similarly to humans. The resulting speech-based interfaces have several advantages over traditional input methods using a mouse or keyboard: to speak is more natural and requires less training, it allows for multitasking, and is also faster than traditional input methods [27].

With the widespread adoption of Natural Language Processing (NLP), today's intelligent ASR systems often use Machine Learning (ML) models for speech recognition. Figure 2.1 shows a commonly used speech recognition architecture. Google's speech recognition system has demonstrated stable recognition accuracy (as low as 8% word error rate) in various environments and domains, and shorter response speed than other systems[9, 28, 34, 10].

The most frequent way to characterize the performance of ASR systems uses the Word Error Rate (WER) [10], which is defined as the number of erroneous words divided by the total number of words. The types of errors are often broken down into insertions, deletions, and substitutions. In the equation below, N is the number of words, and I, D, S represent the number of insertion, deletion, and substitution errors, respectively [10].

$$WER = \frac{I + D + S}{N}$$

Google's ASR system demonstrated an accuracy of 8% WER, which outperformed other state-of-the-art commercial speech recognition systems (such as Microsoft's and Apple's systems, and CMU's Sphinx) in a comparison [10]. Ali et al. investigated [6] the WER per
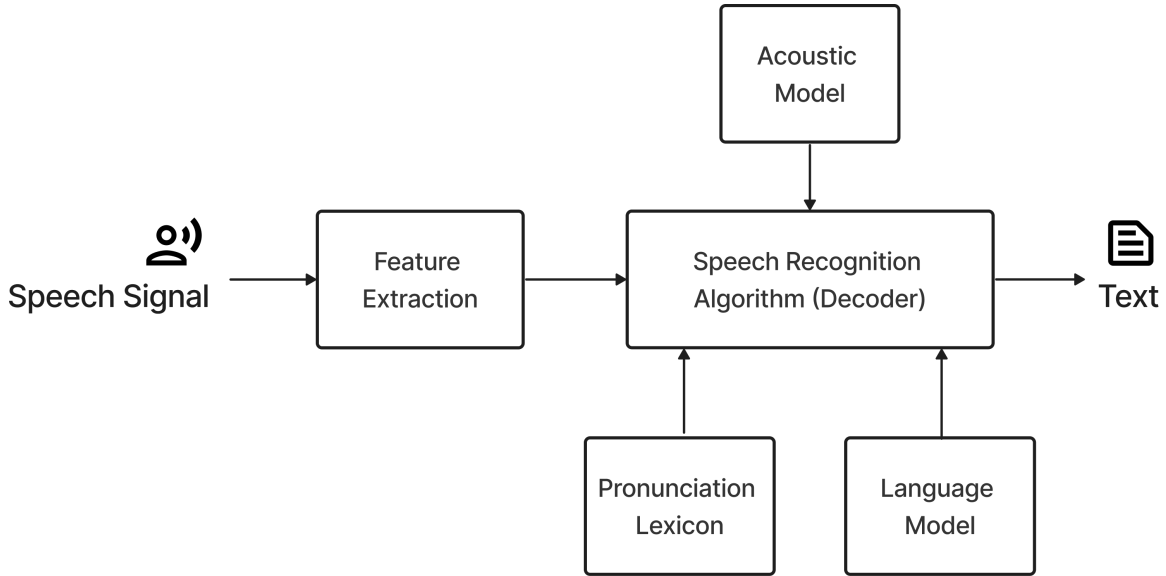
Figure 2.1: Modern speech recognition architecture, after [36, 44].

utterance of two deep neural networks and discovered the WER per sentence to be between 25% and 33%.

2.2 Error correction with voice input

Error correction plays a critical role in text entry. Advances in error correction algorithms have enabled improvements in the efficiency of typing-based text entry and the users' experience [41, 46, 56]. Still, compared to typing, voice-based text input can afford two to five times higher text entry rates [7, 18, 26, 43, 50]. However, the level of acceptance of speech as a means of text input is not as high as typing, one of the major causes being the difficulty to correct text entry errors [30]. Previous research mainly focused on hands-free techniques to correct errors, rather than how users respond to the errors. Such work identified that using multi-modal input, i.e., multiple modes of input, might improve error correction in speech recognition [37, 50]. For example, combining voice input with a mouse or keyboard can make error correction easier compared to using a voice-only approach [23]. In real-world text entry scenarios, there is a tendency to start with voice-based text entry first, and then switch to the keyboard to manually correct the text [20, 30]. This switch of modality can greatly impact the user experience, and is especially challenging for visually impaired users [7]. Azenkot et al. found that visually-impaired participants spent more than 80% of their time editing speech recognition results, with most edits using the backspace key and reentering text with the keyboard [7]. Similarly, sighted people spent 66% of their time editing the text with a desktop dictation system [26].

3

Unfortunately, accurate error correction via voice is more challenging than through typing, due to the linear and temporal nature of audio making it more prone to errors [17]. The core challenge in voice-based error correction is specifying the location and the size of the erroneous (part of the) phrase [11, 26, 37].

To address this issue, McNair and Weibel [33] proposed an effective error correction technique that only requires re-speaking the erroneous part of a sentence (or longer text). With this, users simply re-speak the erroneous word(s), e.g., correcting "I will bang it over tomorrow" by saying only "bring", instead of re-speaking the full sentence. Given that matching the re-spoken part to the original utterance can be challenging and can yield ambiguous results, this technique was further improved by Ghosh et al. [19] (specifically the approach outlined in the appendix only available in the ACM Digital Library). Their approach lets users provide a bit more context by re-speaking more than the erroneous part. The additional words enables the system to detect the incorrect part of the phrase more accurately, making error correction more reliable. In the above example, users could thus, e.g., say "bring it over" to correct the text in an unambiguous manner.

Besides re-speaking, another popular method to correct text using voice is by using the direct command, for instance, replacing a word in the sentence by saying "replacing word x[old word] to word y[new word]". Ghosh et al. compared the two methods and discovered that while commanding worked better with single-word edits, re-speaking performs significantly better with more semantically complex edits [21].

Since I am targeting the error correction of the result of speech recognition for any kind of text, I adopted the re-speaking technique of Ghosh et al. [19] for my evaluation. To ensure that the beginning/end of any speech act (be it the initial utterance or an error correction) was detected reliably, I combined it with minimal use of the mouse. Idiscuss the details of my implementation in Section 3.2.

2.3 Human error correction

According to Zapf & Reason [57], human error handling process involves multiple stages: error detection, error explanation/diagnosis, recovery planning, recovery execution. While error detection refers to user's realization that an error has occurred [57], error explanation refers to knowing how or why the error occurs, and error recovery planning and execution refers to the attempt to undo the error that was made previously [8]. Thus, error recovery is not possible if the error is not detected in the first place. Consequently, error correction techniques are only useful for users after they have detected an error. In text entry with voice input, the error correction methods typically target the error recovery process, but do not facilitate error detection.

Human executive function refers to a family of top-down mental processes that often require effortful concentration [15]. As one of the crucial aspects of this function, error

detection allows me to detect errors and adjust my behaviours appropriately to avoid future errors [47]. Reason [40] suggests three major ways to detect errors, 1) self-monitoring which relies on one's own detection of errors; 2) environmental cues that provide feedback through the environment to identify the errors; 3) identification of errors by other people. In the text entry field, it is common to use environmental cues to help users detect errors, such as the built-in spelling/grammar checkers available in most typing-based tools today. In my work, I rely on the first type of error detection, i.e., self-monitoring in the research design. I made this choice as many speech-interfaces already output only words that occur in the dictionary and/or conform to grammar rules, which renders standard feedback approaches such as autocorrection useless.

Motivation is another factor that can influence error detection. Previous work found participants tend to make fewer errors when they were told that their performance led to reward or punishment, but they were less likely to care about their performance when their performance was not tied to consequences [48].

2.4 Measures of user performance and user experience

To assess participants' interaction with the system, I use two different types of measures: text entry efficiency and subjective experience measures. To quantify entry efficiency, I measure the entry speed, verification time, and accuracy of the entered phrase relative to the prompt. To assess the subjective experience, I measure participants' self-reported mental and physical demand, and their frustration levels. According to the results of previous studies on error correction, I expect to see impaired entry efficiency and an increase in perceived workload and frustration as the number of errors increases [3, 4, 5].

2.4.1 Entry Efficiency

The two most commonly used measures for text-entry efficiency are word per minute (WPM) and error rate (ER). WPM measures the speed of the text entry, with faster WPM indicating more entry efficiency. In voice-based typing, it also refers to how fast users speak. ER measures the accuracy of the text entry, where I compare the given text (in transcription tasks) with the final text users entered, with lower ER indicating more entry efficiency. While Word Error Rate (WER) assesses the error rate on a per-word basis, Error Rate (ER) evaluates the outcome at the phrase level, i.e., it is the ratio of the total number of erroneous trials to the total number of phrases presented. Both WPM and ER have been frequently used in literature to measure text-entry efficiency [1, 3, 5, 7, 26]. I also added another metric used in previous work [4] called verification time (VT) to measure how long it took participants to decide that the result is correct, after they finished entering the text.

2.4.2 Subjective Experience

The NASA TLX is a well-validated questionnaire that was originally designed to measure workload in the military [39]. It has been applied in a variety of settings in HCI research [32]. The NASA TLX combines six different scales including mental demand, physical demand, temporal demand, effort, performance, and frustration. Notably, when digital software behaves unexpectedly, frustration appears to be a common user experience. The unexpected delays or errors that happened in computer systems can cause a high level of frustration among participants, and as a result, impact user performance [24]. Previous text entry research on typing-based text entry systems with autocorrection and prediction identified that errors typically increase the frustration of the users due to the increased mental workload, which also reduced their text entry performance [3, 5, 29, 38, 39].

Chapter 3

Study 1

To understand how users respond to errors encountered in speech transcription tasks, I performed a first study. I used Google's ASR system as my baseline, as it seems to be one of the best-performing systems [10]. Based on the relatively high error rate figures documented in the literature for ASR systems (25%-33% in [6]), related work [35], and my experience from early pilots, I decided to insert errors with a probability of 33% in my first study.

My first study used a between-subject design with the injected error rate (two levels, 0%, and 33%) as the independent factor, with half of the participants being (randomly) assigned to each one of these two conditions. The dependent factors included entry speed (WPM), error rate (ER), as well as the perceived workload. I only measured participants' speaking time to calculate the WPM. The time was calculated from when participants first clicked on recording the button to start speaking, until they clicked the stop button to finish their recording. In other words, I did not include the phrase reading time in this measure. To distinguish user errors from the errors induced in the system, the ER I measure here refers only to user-generated errors, which might stem from failing to correct the text, or issues such as misreading the text, mispronouncing the word, or saying the wrong word by accident. To assess the perceived workload, I borrowed several questions from the NASA TLX for my questionnaire, to measure the frustration and the perceived physical and mental demand. Overall, I collected data for 29 phrases for each participant, for a total of 348 phrases.

3.1 Participants

Twelve participants (six females, six males), aged between 21-29 years old, with an average of 24.5 (SD = 2.15) participated in this study. All participants were either completing or had completed a bachelor's degree in an English-speaking university in Canada. All data were collected over Zoom. Participants generally shared their screen, except for two who did not agree to do so.
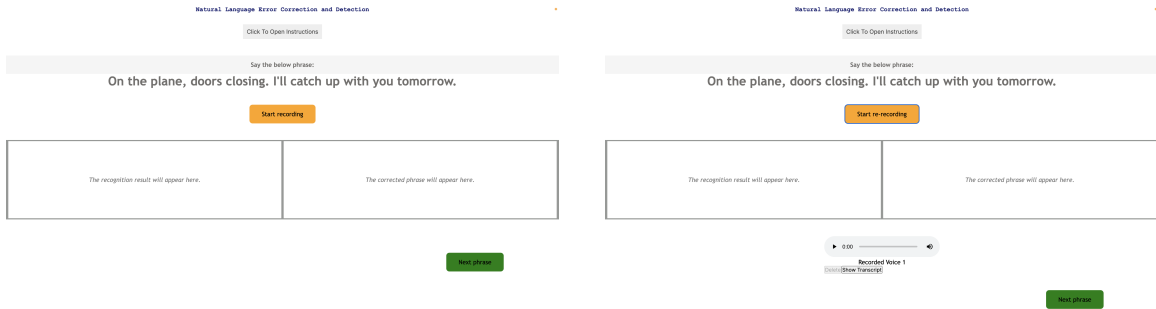
Figure 3.1: Interface of the speech-to-text system in Study 1.

3.2 Apparatus

The experiment used a web application housed on a local university server. I implemented the system using HTML, CSS, JavaScript, and used the Google speech recognition API [22]. Figure 3.1 shows the interface of my system. Normally, such a system would show the most likely recognition result for the user's utterance, but I injected errors (with a controlled frequency) by showing the second-most likely result returned by the Google API, which yields a very believable mis-recognition. For instance, instead of showing "How was your trip to Florida?" (the most likely result), the system would then display "How was your train to Florida?" (the second-most likely result). In the two experimental conditions, I injected such an error either 0% or 33% of the time. I chose 33% in this study as a compromise between being able to observe sufficient errors and avoiding excessive user frustration. For error-correction through the re-speaking feature, I enabled two attempts for transcribing each phrase for the participants. For the first attempt, users were required to speak the whole phrase in full, while the second attempt was reserved for correcting any potential errors that remained after the first attempt. Errors were only injected in the first attempt, and the system always showed the most likely result in the second attempt.

3.3 Task and Procedure

All phrases were randomly selected from the Enron MobileEmail phrase set (see Table 3.1 for example phrases) [54]. To alleviate the mental burden on participants and to simplify the data analysis, I removed all punctuation marks in the recognition results, similar to other work on text entry [4, 5, 49].

At the start of this study, and as supported by the version of the Google ASR system available at that time, participants were allowed to choose the most appropriate accent, among English-US, English-UK, English-India, and English-Canada. Participants were then

I will bring him. He would love anything about rocks.
Hope your trip to Florida was good.
She called and wants to come over this AM.
There is now a meeting at 8PM as well. See you soon!
It reads like she is in.
Has he made a specific request?
I am walking in now. Tell her to get my expense report done.
I am out of town on business tonight. You must wait until she comes home.
Not even close. They are more efficiently pooled.
Could you try ringing her? This message concerns me.
Do you need it today? Keep me posted!
Call me to give me a heads up.
What is in the plan?
Where do you want to meet to walk over there?
I am almost speechless. Please put me on the list.
Suggest you get facts before judging anyone.
Have I mentioned how much I love this city's traffic?
Should systems manage the migration?
Take what you can get. Take what you can get.
I have never worked with her. Get in touch with the accountant instead for format.
I am glad you liked it. Thank you.
I hope you are feeling better.

Table 3.1: Example phrases used in Study 1.

asked to speak each phrase that appeared on the screen. Ten participants chose the US accent and two the Canadian accent.

Participants clicked on "Start Recording" to start recording their utterance (Figure 3.1 left), followed by a click on "Stop Recording". If the speech was transcribed incorrectly, then they could repeat part of the sentence by clicking on the "Start re-recording" button on the same page (Figure 3.1 right), and then again "Stop Recording".

Instruction on the Re-speaking Feature. I asked participants to (generally) re-speak the incorrect phrase starting from at least one word before the incorrect word and ending at least one word after the incorrect word. If the correction involved the first or last word of the target phrase, then they were asked to repeat two words after or before. The system did not enforce this instruction, instead the algorithm matched whatever participants used to correct an error. Participants were only given a single error correction attempt for each phrase, after which they had to proceed to the next phrase.

After participants completed all 29 phrases, they completed a NASA TLX 7-point Likert scale questionnaire (with 1 = lowest, 4 = neutral, and 7 = highest). Each participant was asked to self-report their subjective experience, followed by a brief interview at the end to assess participants' familiarity with speech-to-text systems and their experience with the re-speaking interface.

3.4 Quantitative Results

To analyze the error rate, I compared participants' final submitted transcription results with the phrases they were given. Three types of errors could occur in my study: naturally-occurring errors, the errors I injected, and errors that resulted from the occasional glitch in my web server and/or the speech recognition system. For my study, I am specifically interested in the first type of errors — naturally-occurring errors that are caused by the insertion of system errors (i.e., the second type).

Figure 3.2 shows the results for all dependent variables. Overall, error rates for participants in the 33% error condition (M = 29.9%, SD = 13.7%) were higher than with 0% errors (M = 24.1%, SD = 5.8%). However, there was no significant difference between the two conditions, $t(6.71) = 0.94$, $p = .37$. See Figure 3.2a.

Also, participants achieved a higher WPM measure in the 0% error condition (M = 122.37, SD = 28.66) compared to 33% errors (M = 118.10, SD = 38.15). Yet, a Wilcoxon test revealed no significant difference for entry speed ($Z = 1.58$, $p = .11$). See Figure 3.2b.

In the 0% error condition (M = 1.50, SD = .55) participants reported lower physical demand than with 33% errors (M = 1.67, SD = 1.63), but the difference was not significantly different ($Z = - .76$, $p = .44$). See Figure 3.2c.
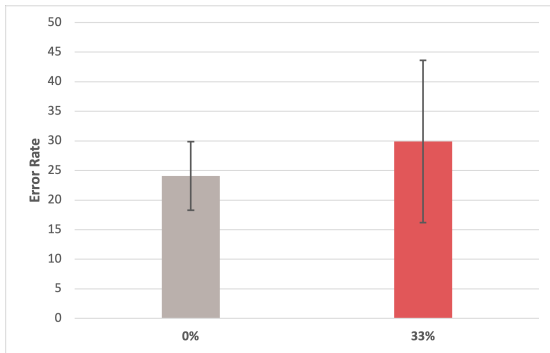
(a) ER
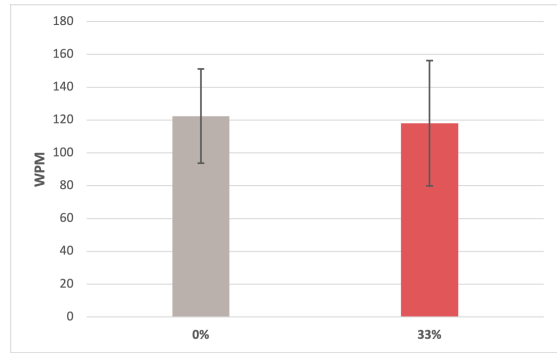
(b) WPM
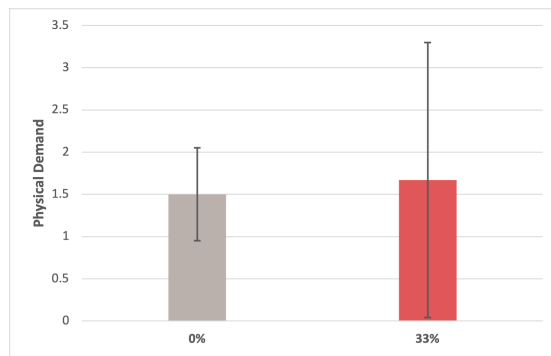
(c) Physical Demand
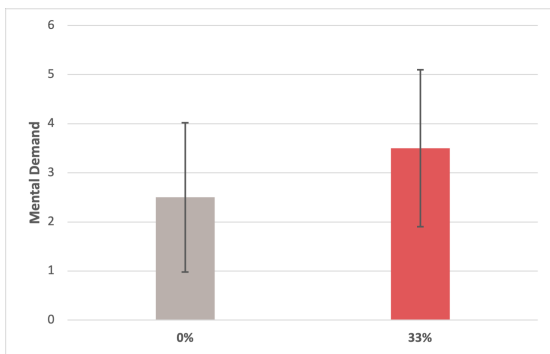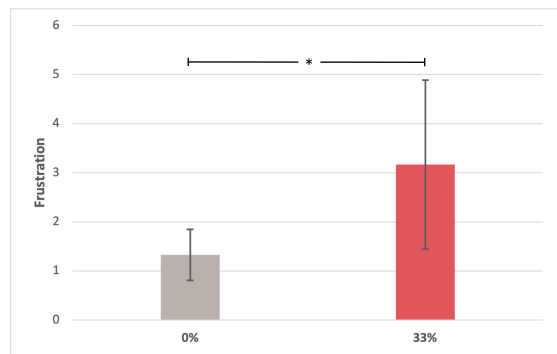
(d) Mental Demand

(e) Frustration

Figure 3.2: Average results of Study 1: a) error rate (ER), b) words per minute (WPM), c) physical demand, d) mental demand, and e) frustration. The error bars represent the standard deviation and horizontal bars illustrate significant differences.

The mental demand for participants in the 33% error condition (M = 3.50, SD = 1.60) was higher than with 0% (M = 2.50, SD = 1.52). However, the result was not significantly different, t(0.74) = 1.02, p = .48. See Figure 3.2d.

Overall, frustration for participants in 33% error condition (M = 3.17, SD = 1.72) was higher than in the 0% condition (M = 1.33, SD = .52). A Wilcoxon signed-rank test between the 0% and 33% error conditions revealed a significant difference for frustration (Z = 1.95, p = .05), with a small effect size (η_p^2 = .03). See Figure 3.2e.

3.5 Qualitative Results

At the end of the study, I conducted a brief, semi-structured interview with each participant to assess their familiarity with voice-to-text systems and their general experience with my system in comparison to other voice-to-text systems they had used before. All participants reported at least some level of familiarity with speech-to-text tools during my interview. When I asked about their experience with my system compared to other systems they had used before, participants had very diverse responses, regardless of which condition they experienced. For example, a participant in the control condition reported the system as *"very dumb, needs more development"*, whereas another participant in the experimental condition said: *"It has better accuracy, understood me better."* When I asked them about the most challenging part of the study, a majority–7 out of 12 participants–reported disliking the phrases I used in the study. Some said the phrases felt weird and unnatural, took them a long time to read, and they also made more mistakes. Others mentioned that some of the phrases were too long.

3.6 Discussion

Even though I induced a non-trivial amount (33%) of errors, one of the most striking findings from my study was the lack of a significant difference. This contrasts starkly with previous work investigating the effect of autocorrection failures in *typing-based text input*, where an (inserted) error rate of 10% and 20% already yielded a large effect on users' typing performance and perceived workload [4]. While I found a significant effect for frustration, the effect size was small in this study, which indicates that the difference between the two groups might not be that substantial. Yet, one of the key takeaways from the qualitative results was the apparent inappropriateness of the phrase set for a task that involves speaking. I had extracted that phrase set from a written corpus. Yet, based on my participants' responses, I believe that the properties of that phrase set could be one of the most likely explanations for my results.

Chapter 4

A New Phrase Set for Speech Recognition Evaluations

To address the issue identified with the phrase set, I decided to create a new phrase set more appropriate for the evaluation of speech recognition systems.

4.1   Spoken vs. Written Language

Most current text entry studies employ transcription to evaluate text entry efficiency. Yet, the most widely used phrase sets in such studies were all designed for typing, not for speaking [32, 52, 54]. Still, relative to text transcription, text composition is more representative of real-world text entry scenarios. Although the composition task has higher external validity than the transcription task, the latter is often more appropriate in experimental evaluations due to its higher internal validity and lower variability [55]. A recent study [18] let participants compose their own phrases, and use them for transcription tasks with other participants, which increased the logistic effort substantially. According to Foley et al. [18], phrases for transcription tasks have to fulfill the following criteria: memorable (users can enter a phrase after the prompt without referring to it), representative (resemblance of the actual text to text entered by people), and replicable (the phrase set is publicly available). For my work, I am specifically interested in language production for speech recognition, not language interpretation.

Based on the results of my first study reported above, it seems that a set of phrases based on written English might not be suitable for speech-based text entry evaluation. However, given the small sample size in Study 1, I could not tell whether the phrase set is an obscuring factor or a confound. Yet, I believe that using such an un-representative phrase set might have introduced substantial noise into my study. Spoken and written language contrasts in various aspects: spoken language is less abstract, has more finite verbs, and has fewer nouns of abstraction. There is also a contrast in syntax and sentence structure, and in terms of the manner and speed of production [1]. Moreover, entering text by tying versus speaking can

The auto industry is going to have to innovate here a little bit.
There are huge long waiting lists for classes.
Why don't we get more of that rewards?
It's hard for me to understand, but it happens, and that is what the Bible teaches.
But then maybe she needs to take the whole family and leave and never come back.
We've talked to other people like Julia, people who don't normally come out in protest for various events.
What are the risks versus the rewards?
You do not have to have every single piece of the puzzle to determine the picture.
But it was extraordinary to find somebody like that as an eyewitness to such an event.
I don't think that I've seen as many of you in one place since they announced the results of my first bar exam.
Drug abuse is now the number one cause of accidental death, according to the federal government.
Well, this is like a chess game I have to anticipate ten steps ahead of him.
I don't think it will come as any great surprise to you that I stopped reading newspapers.
I don't think that this set Donald Trump off his game or anything like that.
I started my own first band when I was 45.

Table 4.1: Example phrases from the new phrase set used in Study 2.

lead to very different experiences for users [17, 43]. Neuroscience research also identified that written and spoken language involves two distinct systems that are controlled by different parts of the brain [39]. Therefore, phrase sets for spoken and written language are likely not interchangeable. In psychological research, error detection involving different sensory systems, such as reading, typing, and speaking, is also often studied separately [31]. In my work, I am specifically concerned with error detection in reading (reviewing the text) and speaking (transcribing the prompt), not hearing or understanding.

4.2   Phrase set for speech-based text entry

To address the issue identified above, I collected a new phrase set (see Table 4.1 for example phrases) from spoken English corpora. Consequently, I selected my phrases only from spoken English corpora, using the Santa Barbara Corpus of Spoken American English (SBCSAE) [16] and https://www.english-corpora.org [14], unlike the majority of text-entry research, where the phrases were (largely) selected from written text, such as the Enron MobileEmail phrase set [42].

Given the many potential uses of speech recognition in everyday scenarios, the main focus of my research is to investigate speech recognition for transcribing informal conversations. It is worth noting that the causal conversational use of speech recognition can differ from professional context/high-stakes conversations, where the latter has a higher resem-

blance to written text. To increase external validity, I decided to focus on the more common use case of informal conversations.

I initially selected 1000 phrases from the spoken English corpora, but removed all those that contain inappropriate language, uncommon names, or complex vocabulary, to avoid potential confounds within speech-based text entry evaluations. I ended up with 768 phrases, all with generally short to medium length, with on average 11.5 words (SD 4.14, ranging from 6 to 27). See the Appendix for the full phrase set.

Chapter 5

Study 2: Re-evaluation of the System Using the New Phrase Set

To address the issue of generally negative opinions on the phrases used in Study 1, I conducted a second study with the new phrase set and the improved version of my web system and its user interface. As I had identified a fairly high tolerance to errors in the transcription results in my first study, I decided to investigate an even higher error rate, 50%, in this study. Based on the findings from the first study I also re-specified my research questions for this study to be:

RQ1. How do errors injected in speech transcription tasks impact users' entry speed, error rate, verification time, and perceived workload?

RQ2. Do different levels of error injected impact users' responses differently?

I reused all dependent variables from the first study, but added a measure of Verification Time (VT), which refers to the time that participants took to review a phrase before moving on to the next one.

5.1   Study Design

In this study I used a within-subject study design, with every participant participating in all three conditions, which only varied in terms of the frequency of injected errors. I injected an error 0% (control condition), 25%, or 50% of the time in the three conditions. The order of the conditions was counterbalanced among participants. I also recorded experimental observations during the study and performed a semi-structured interview with each participant at the end of the experiment. Although such qualitative responses were not the main focus of my study, I included them as they might provide valuable insights about participants' perceptions and inspire future work.

Natural Language Error Correction and Detection

Participant ID

Please enter your Participant ID

Choose Your English Accent

US

Begin study

Please speak the phrase below:

I knew that you would come to your senses.

Start recording   Complete study

| *The 1st recognition result will appear here.* | *The corrected result will appear here.* |

Next phrase

Figure 5.1: Interface of the speech-to-text system used in Study 2.

Figure 5.2: All sixteen options for English accents I offered in the system (supported by Google).

5.2 Apparatus

I adapted the system used in Study 1, by improving user interface (see Figure 5.1), and also adding support for all sixteen English accents supported by then current version of the Google Speech-to-Text API (see Figure 5.2). I also included the option for participants to either use the mouse/trackpad or keys on the keyboard to control the recording/next buttons (space bar for start/stop recording and right-arrow key for the next phrase).

5.3 Participants

I recruited eighteen participants from the local university (mean age = 23 years, SD 5.12, 15 female) via on-campus flyers and word-of-mouth. All participants reported intermediate to professional levels of English-speaking skills. Most participants had had some experience

with speech-to-text systems, and only one reported that they had never used such systems before. Ten participants reported English was not their first language. Overall, nine participants chose the US accent, eight the Canadian accent, and one the Indian accent.

5.4   Procedure

After indicating their informed consent, I gave participants detailed verbal instructions for the task, followed by a brief demonstration of how the system works. Participants then were given practice trials until they demonstrated an adequate understanding of the system, which all participants were able to do within five phrases. Due to the limitation of the Google Speech-to-Text API, participants were told to ignore all punctuation marks, contraction issues, and spelling inconsistencies such as Canadian versus US spelling. Each participant entered a block of 30 phrases for each of the three conditions. Thus, each participant completed 90 phrases. All participants experienced the phrases in random order, with the order of conditions counterbalanced across participants. In total, I recorded 90 x 18 = 1620 phrases.

At the start of the experiment, participants were asked to choose their preferred English accent among the sixteen English-speaking countries and regions supported by the Google Speech-to-Text API. As in Study 1, participants were then asked to speak each phrase that appeared on the screen in the prompt area. Participants clicked on "Start Recording" to start recording their utterance, followed by a click on "Stop Recording". If the speech was transcribed incorrectly, then they could repeat part of the sentence by clicking on the "Start re-recording" button on the same page (Figure 5.1), and then again "Stop Recording".

I asked participants to (generally) re-speak the incorrect phrase from at least one word before the incorrect word and end at least one word after the incorrect word. If the correction involved the ending or starting word, then they were asked to repeat two words after or before. The system did not enforce this suggestion, instead the algorithm matched whatever part of the text participants corrected. Participants were only given a single error correction attempt for each phrase, after which they had to proceed to the next phrase. To avoid increasing frustration unduly, no errors were injected in the correction attempt.

Immediately after each block, I asked participants to fill out a NASA TLX questionnaire, assessing their perceived workload in the respective block. After they completed all three blocks, I conducted a short, semi-structured interview with participants regarding their overall experience with the system.

5.5   Results

In this section, I summarize the results of my second user study. I only list results that are statistically significant or are worth reporting because they are unexpected, e.g., because they differ from previous work. I first removed all log entries where the system had not

worked correctly due to a small implementation issue (less than .03% of the data) and then used JMP 17 and SPSS for all analyses. Since both the raw and log-transformed data did not fulfill the normality assumptions for parametric tests, I use Friedman tests with post-hoc pairwise tests (Bonferroni-corrected) for all quantitative analyses. I use Kendall's W to calculate the effect size. Below, I first examine participants' overall performance in the three conditions in terms of entry speed, verification time, and error rate. Beyond this, I also performed multivariate analyses on all measured variables, but saw no substantial differences from the corresponding Friedman tests. Additionally, there was no effect of the accent on the results (p > .05). Further, I observed no order effects in all variables I measured, which confirms that the counter-balancing had worked as intended. I then look at the participants' subjective experience of the system based on the responses to the questions from the NASA TLX. Finally, I review the qualitative observation and interview results.

5.6   Entry Efficiency

As mentioned above, I used the word-per-minute metric (WPM) to measure participants' entry speed. Speed was calculated based on the number of words participants entered and their total completion time for each trial. I did not find significant differences across conditions, $X^2(2) = 5.17$, p = .07. See Figure 5.3a.

For the analysis of the remaining error rate, I disregarded spelling variations due to different English spellings, such as "honour" versus "honor"; contractions such as "we're" versus "we are"; and common word combination issues such as "on to" versus "onto". The results indicated a significant difference across the three conditions, $X^2(2) = 4.85$, $p < .001$, with a medium effect size, $W = .45$, p < .001. A post-hoc test revealed significant differences between all pairs of comparisons (p < .001), with a mean error rate of .17 (SD .09) in the 0% error condition, which was lower than in both the 25% condition (.24, SD .06) and the 50% condition (.29, SD .11), both of which were also significantly different. See Figure 5.3b.

For measuring the verification time (VT), I logged three VTs, with VT1 indicating the time elapsed from when participants saw their first recognition result and when they clicked either the "Start Re-recording" or the "Next Phrase" button; VT2 indicating the time elapsed from when participants saw the phrase resulting from the second speech recognition input, i.e., the error correction, and when they clicked on the "Next Phrase" button; and the total VT that adds these two times together. All three VTs were measured in seconds. The results indicate that there were no significant differences across all conditions for VT1 (p > .05, see Figure 5.3c). Figure 5.3d illustrates the statistically significant difference for VT2, $X^2(2) = 45.86$, $p < .001$, with a small effect size, $W = .04$, p < .001. Post-hoc test revealed significant differences between the 0% (1.20, SD 2.80) and 25% condition (1.75, SD 2.84, p = .003), and the 0% error (1.20, SD 2.80) and 50% condition (1.96, SD 2.75, p < .001), but not between the two conditions where errors were injected (p = .13). Figure 5.3e illustrates

20
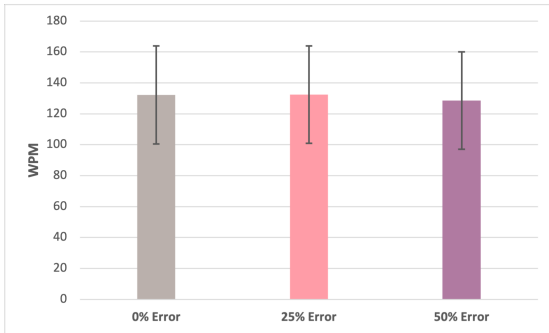
(a) WPM
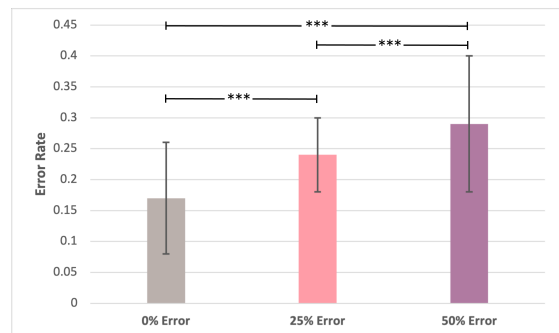
(b) ER
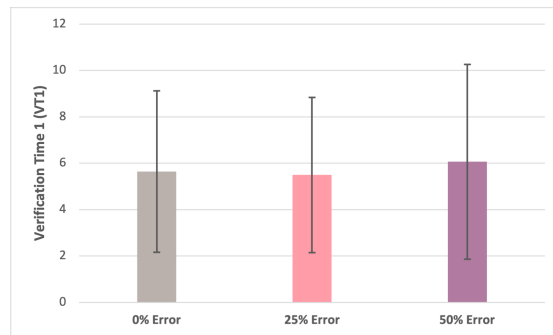
(c) VT1
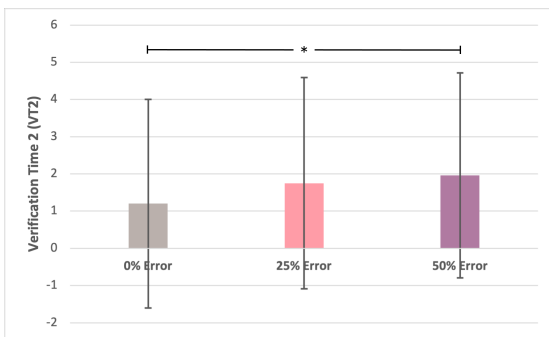
(d) VT2
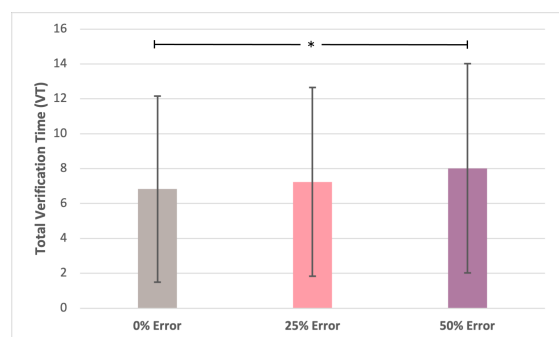
(e) VT

Figure 5.3: Results of Study 2: a) average words per minute (WPM), b) average error rate (ER), c) first verification time in seconds (VT1), d) second verification time in seconds (VT2), and e) total verification time in seconds (VT). The error bars represent the standard deviation.

(a) Mental Demand

(b) Physical Demand
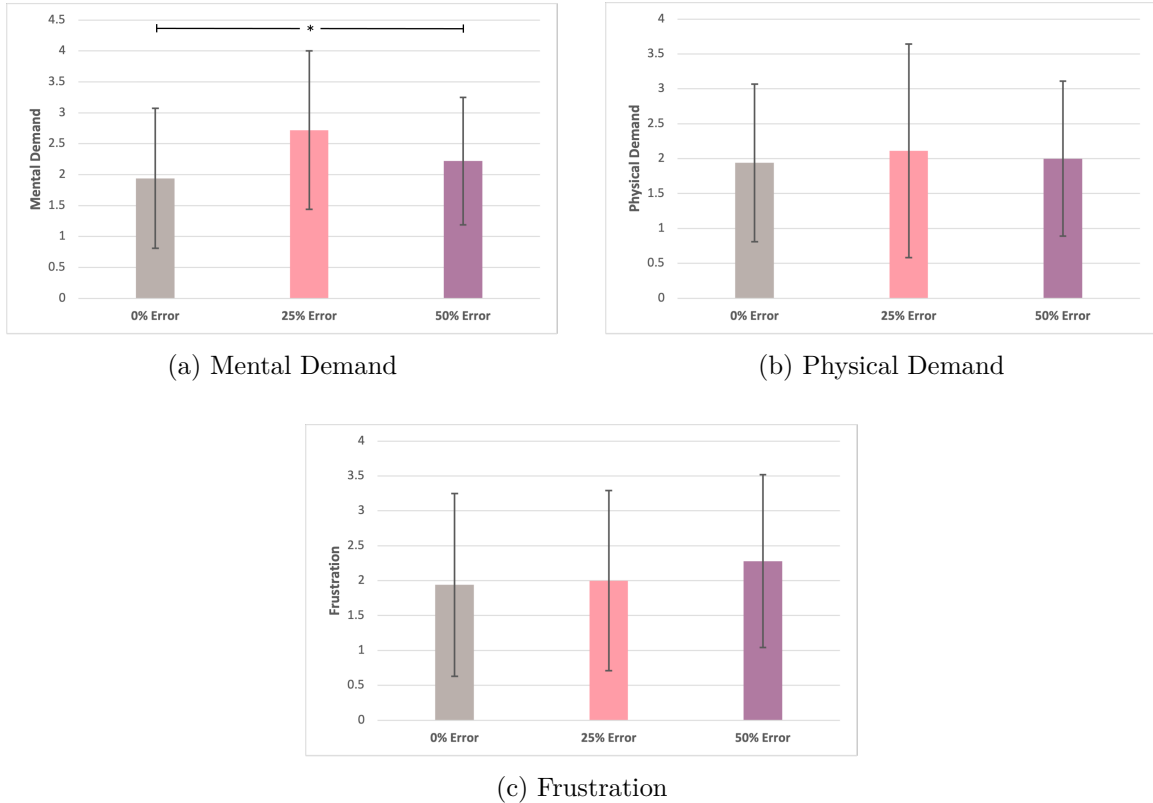
(c) Frustration

Figure 5.4: Results of Study 2: NASA TLX scores for a) average mental demand, b) average physical demand, c) average frustration. The error bars represent the standard deviation.

the statistically significant difference for VT $X^2(2) = 14.27$, $p < .001$, with a small effect size, $W = .01$, p $< .001$. Post-hoc test revealed a significant difference only between the 0% (6.84, SD 5.33) and 50% (8.02, SD 5.41) conditions.

5.7 Subjective Experience

To assess the participants' subjective experience, I asked participants to rate their perceived workload (NASA TLX) for the three conditions on a 7-point Likert scale (with 1 = lowest, 4 = neutral, and 7 = highest). The statistical analysis identified a significant difference in participants' self-reported mental demand across conditions, $X^2(2) = 8.27$, $p = .02$, with small effect size, $W = .23$, p $= .02$, see Figure 5.4a. Yet, a post-hoc test revealed no differences between all conditions (p $> .05$), suggesting a spurious result. There were no significant differences for physical demand (Figure 5.4b) nor frustration (Figure 5.4c), all p $> .05$.

5.8 Identification Errors

To further explore the user's behaviour, I examined the different error outcomes in the final submitted phrases. Based on the final submitted text, there were four different categories of outcomes, see Figure 5.6 for an overview. The text could be either correct (indicated by the blue hues) or incorrect (indicated by the orange hues). Participants could have also either got the phrase correct in the first attempt (indicated by the dark tones) or used the correction attempt (indicated by the light tones), regardless of whether the submitted phrase was correct or not. I noticed that participants used only the first attempt in the 0% error condition 76.4% of the time, in contrast to 62.3% in the 25% error condition and 53.9% in the 50% error condition, which was consistent with my error intervention, i.e., I expected more error corrections to happen as the number of injected errors increased.

Yet, when the final submitted phrase was incorrect, using only the first attempt indicates a failure of the user to identify errors in the first recognition result (i.e.,"Identification Errors", see Figure 5.6, indicated by the orange arrow). The identification error rate was 5.6% in the 0% error condition, followed by 6.5% in the 25% condition, and 8.1% in the 50% error condition. Such identification errors account for 6.7% of all submitted outcomes, and 28.5% of all final submitted erroneous phrases. Hence, I observed a potential trend between the induced error rate and the detection of such errors. In other words, as the reliability of the system decreases, participants were less likely to detect such errors. There is a moderate correlation between injected error rate and the identification rate (see Figure 5.5), suggesting a moderate portion of the variability in the identification error can be explained by the injected error.

As mentioned in the related work section, among the three ways to identify errors (by themselves, due to the environment, or by others) participants had to solely rely on their self-monitoring [40, 47] in my study to identify errors, since there was no indication for errors by the system nor an intervention by the experimenter. Also, since my participants were not informed how well they performed nor faced consequences for inadequate performance, they might not have had enough motivation to correct more errors [48].

5.9 Time Series Analysis

Considering that my system was novel for the participants, I also examined my data for potential learning effects. For this, I ran an analysis over the time series across the number of phrases. Yet, I saw no strong indication that the number of phrases had an impact on users' performance in terms of VT (see Figure 5.7) and WPM (see Figure 5.8). I also did not observe learning or fatigue effects, suggesting that the number of phrases was too small to elicit such responses.
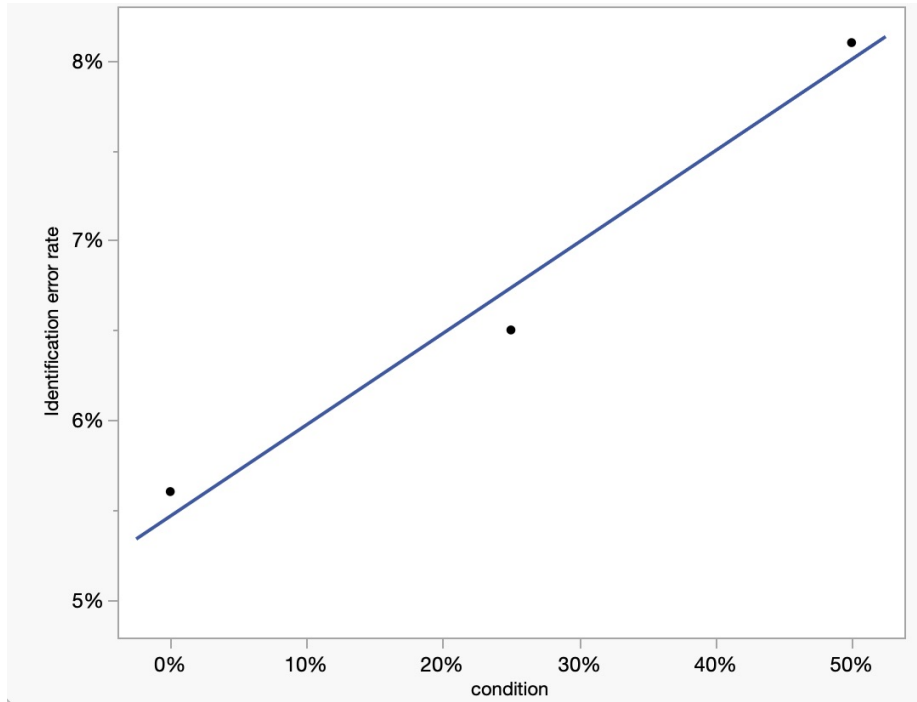
Figure 5.5: Correlation between injected error rate and identification error rate. The x-axis represents the injected error rate (from 0% to 50%).
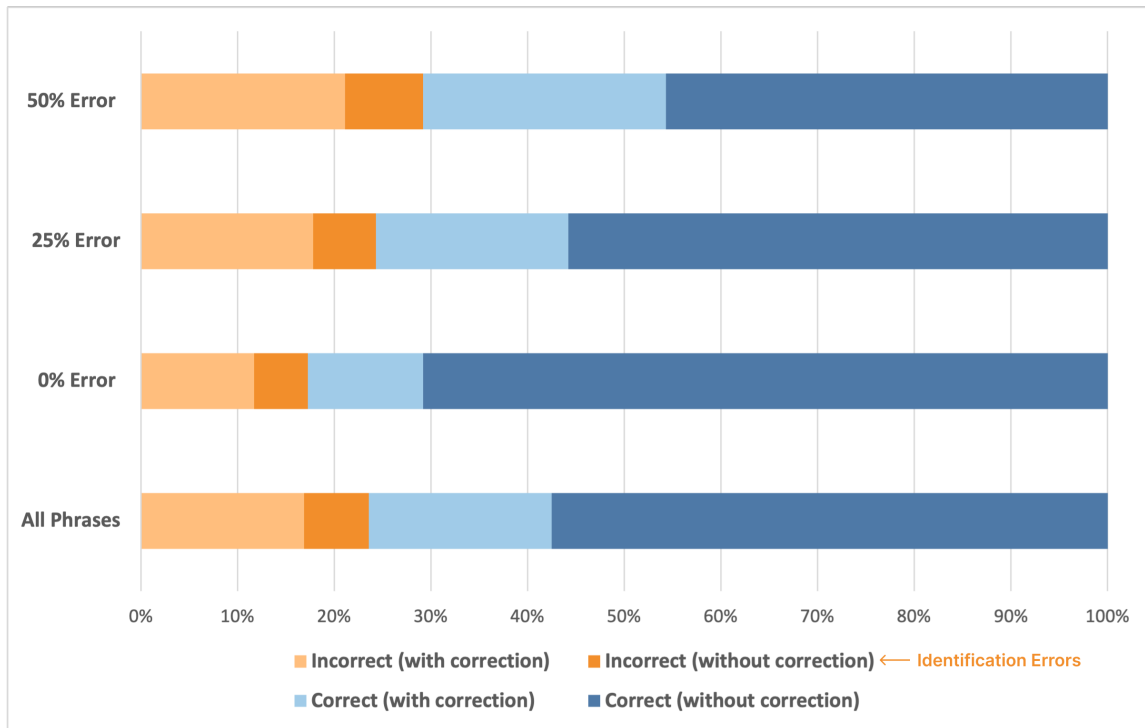
Figure 5.6: Possible outcomes in the final submitted phrases across all conditions.
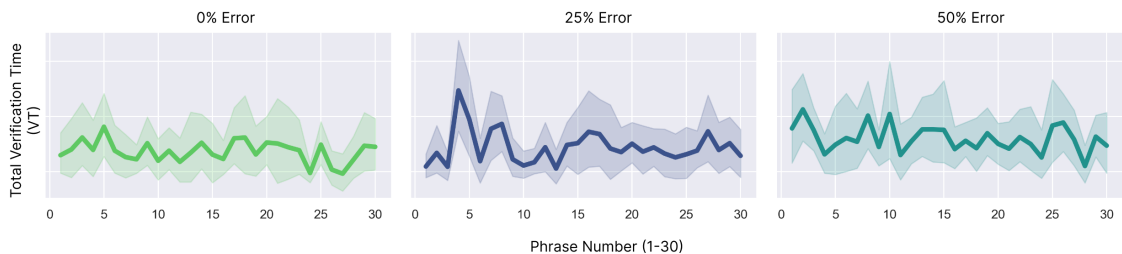
Figure 5.7: Effect of the phrase number on the verification time, the x-axis shows phrase 1 to 30.
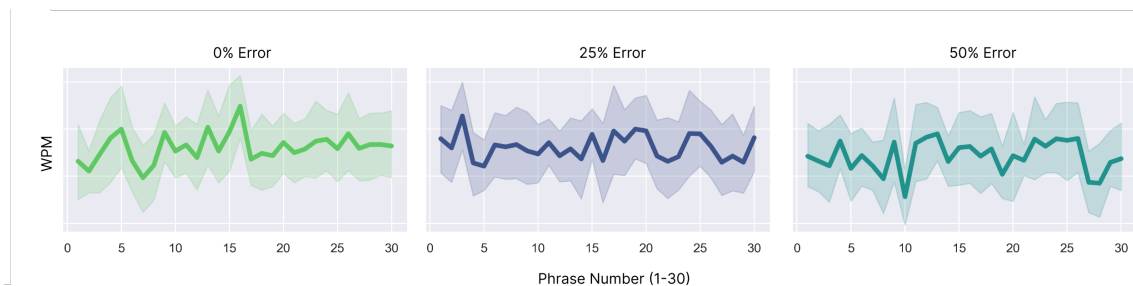
Figure 5.8: Effect of the phrase number on text entry speed, the x-axis shows phrase 1 to 30.

5.10 Evaluation of the Phrase Set

To evaluate the effect of the phrase set, I compared the error rate between Study 1 and Study 2, see Figure 5.9. For this, I ran a Welch's t-test for analysis. I discovered that participants made significantly fewer errors in Study 2 (M = 17.3%, SD = 9.7%) than in Study 1 (M = 24%, SD = 5.6%), t(15.5) = -2.07, p = .03, with Cohen's d = .75, indicating a large effect of the phrase set on ER. I also compared the WPM between Study 1 and Study 2 and discovered that participants were significantly faster in Study 2 (M = 132.27, SD = 31.69) than Study 1 (M = 122.37, SD = 28.66), t(320.4) = 3.86, p < .001, with Cohen's d = -.32, indicating a moderate effect size for the phrase set on WPM. Together, these findings suggest that my new phrase set did lower error rates and increased participants' entry speed when there were no errors inserted in the task. For reference, the speaking rate for English conversation typically falls between 130 to 220 WPM [51]. Yet, since these speech rates are obtained for productive speech, I can expect the rate to be at the lower end for speech transcription tasks.

(a) ER Comparison   (b) WPM Comparison
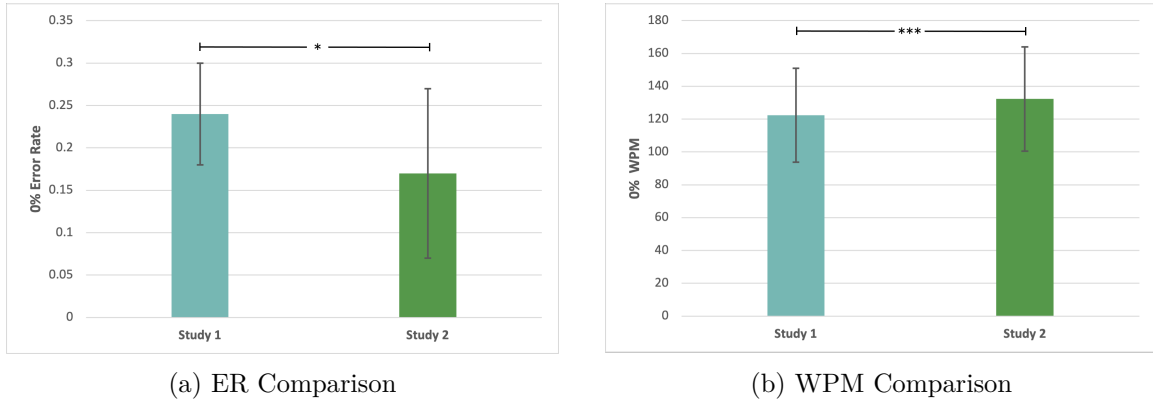
Figure 5.9: Comparison of a) Error Rate, and b) WPM between Study 1 and Study 2 in the no error condition. The error bars represent the standard deviation.

5.11 Qualitative Observations

Although the Google Speech-to-Text API supports sixteen different English accents, 15 out of 18 participants selected American or Canadian English, and only three participants chose to actively look for the variant that corresponded to their native language. In the end, only a single participant selected a non-North-American accent (Indian accent). The other two participants stated that they did not find their preferred accent listed as an option. Some participants also commented that they did not know whether they had an accent or not.

During the task, I observed that participants tended to review the given phrase before they started speaking it, but this only occurred when they encountered longer phrases. Also, and corroborating the results on identification errors discussed above, it was not uncommon for participants to overlook errors in their recognition results, especially for shorter words such as "and" or "the", especially when transcribing longer phrases. For such small words, they occasionally failed to notice errors related to them and then proceeded to the next phrase, not even using the second step, i.e., the correction attempt.

Although I provided the option to control the buttons using either keyboard or mouse/trackpad, I observed a strong preference for the mouse.

5.12 Questionnaire and Semi-Structured Interview Results

Overall, participants rated their experience with the system between neutral to positive. More than half of the participants directly mentioned that they liked the re-speaking feature, as it was convenient and time-saving. Since 17 out of 18 participants reported occasional usage of speech-to-text systems, I focused on the comparative experience of my system with other systems they had used in the past. When I asked them to name a few systems they have used, the most frequent answers were iOS built-in dictation, Google's pages/apps (Translation, Search, and Google Home), and Microsoft Word. There was a tendency to

use such systems only when their hands were busy, such as while driving or cooking. Since the majority of the participants were university students, using such systems for lecture transcription or essay writing was also a common answer. For social purposes such as texting, there was a clear preference for typing, as they perceived it to be more accurate. In terms of accuracy, 11 out of 18 participants commented positively on the accuracy of my system, and 4 of them even thought the system was more accurate than the speech-to-text systems they had used in the past.

On the other hand, I also received various forms of negative feedback. For example, participant 2 said *"It was hard that I had to only use voice to correct the sentences, I wish I could type."* Participant 17 said *"When I have a longer pause, the system won't pick up properly, or add extra words to my sentences. When there was slang in the sentence, it keeps changing to formal English. For example, I say 'gonna' but the system keeps recognizing it as 'going'. I prefer a more slang-friendly system."* Two participants also mentioned that the re-speaking feature did not work well when there were two incorrect parts in the first recognition result (which was indeed a limitation of my implementation, as I supported only a single correction "action").

I also asked participants what they thought the most challenging part of the task was, and two participants responded that it was challenging for them to get used to the re-speaking feature. Other participants had more diverse responses, for example, participant 3 said: *"When sentences were long, I had to read it before I started recording. If I didn't go through it first, I might not pronounce some of the words very well, especially since English isn't my first language."* Participant 9 said, *"I had to make sure I speak really slowly and clearly for the system to work. This was the first time I use a speech-to-text system, and I don't see the need to use it in the future."* Some participants also mentioned that sometimes the re-speaking feature failed to correct the phrase as they intended, e.g., due to a rare matching failure or because the user actually used the wrong words in their correction attempt. Some participants also found that it was difficult to use the re-speaking feature when there was *more than one* erroneous part in the first recognition result, especially when the erroneous parts were not adjacent, e.g., one error at the beginning and the other at the end of the phrase. Beyond the limitation to a single correction action mentioned above, (rarely occurring) failures of my web system sometimes duplicated text rather than replacing it; added extra words to the phrase; or failed to correct the phrase when the erroneous part was at the very beginning of the phrase. After manually screening all the log data for such instances, I concluded that all such system failures had occurred in less than 1% of all trials.

Participant 18 said *"When there was an error, I first had to pinpoint the error and then changed my speech style for it to recognize. But this is just how I talk, I don't know what to change. I feel like technology should adapt to us, instead of the other way around."* Some participants also mentioned the difficulty of getting homonyms (words with the same

pronunciation but different meanings) to be recognized correctly, such as "sight" versus "site".

Additionally, I asked participants if there are any other ways they could think of to edit the phrases using voice. Two participants said they would rather just re-speak the entire phrase, instead of having to locate the erroneous part and correct it. Four participants directly mentioned using voice instruction to correct the phrase, and participant 5 suggested: *"It might be better to have something like the 'Hey Siri feature', you can then just say something like 'Correct this word to that.'"* Two participants stated that they would like to have a third attempt to correct the phrases, *"Especially when there are two incorrect parts"* (participant 15). Participant 11 suggested that: *"Maybe you can say a keyword to put you in the error correction mode. For example, you say 'correction [keyword]: A [wrong word] to B [correct word] '"*

In contrast to Study 1, I did not receive any negative feedback regarding the phrase set. Some participants commented that the presented phrases felt often more conversational than the transcribed result (also because the speech recognition changed their recognition result to more formal English).

5.13 Limitations

One of the major limitations of my study is a potentially reduced external validity. Although a transcription task has high internal validity and enables a clear definition of error rates, people usually compose original text in real-life scenarios. In addition, participants' familiarity with the language and speaking habits can also vary widely, which can also affect the internal validity [55]. Large-scale studies by Cuskley et al. found that native and non-native English speakers can differ significantly in their language use, especially in the use of verb forms [12]. Another potential limitation concerns how time was measured in this study. As mentioned previously, time was calculated from when participants first clicked on recording the button to start speaking, until they clicked the stop button to finish their recording. However, since participants might not have started speaking immediately after they clicked on the recording button, this measure might be slightly inaccurate, which could be improved in my future study. Although all my participants appeared to be fluent English speakers, most of them did not have English as their first language. Thus, certain phrases and jargon that were common for native speakers might not be as natural for non-native speakers.

Chapter 6

Discussion and Future Directions

Among all conditions and all text entry efficiency metrics I measured, I only observed a significant difference with a large effect size on the error rate, with 0% error condition producing the lowest amount of errors, relative to the two other conditions. It seems obvious that the final text tended to contain more mistakes when more errors were inserted by the system. Since I only inserted errors during the first attempt, I can conclude that, as system-generated errors increase, a re-speaking feature is not a sufficient mechanism for full error recovery.

Still, a larger percentage of injected errors in the system did not lead to longer verification times or a slower entry speed. Although I discovered a significant difference in the second verification time (VT2) between the 0% error and 50% error conditions, the effect size was small, which might indicate a potentially spurious result. The increased second verification time might also indicate that users were more careful during the text correction phase when a larger number of errors were present. Given these results, the most surprising outcome of my study is that even though I injected a large amount of errors, 50%, i.e., they encountered on average an error in every second phrase, users did in general not perform substantially differently with the system. Nonetheless, I discovered that as more errors were present in the system, participants tended to detect them less often, causing an increase in the identification error rate of 0.9% and 2.5% in the 25% and 50% error conditions, respectively, relative to the no error condition. Unfortunately, I do not yet have sufficient evidence to explain why such identification errors occur. In regards to the NASA TLX questions, I only observed a significant difference in mental demand between the 0% error and 50% conditions. Still, the effect size was rather small for this result, too. Further, participants' ratings tended to be generally low in terms of the NASA TLX scores.

Further, I also found a positive correlation between the length of the given phrase and WPM, $r(425.23) = .46$, $p < .002$, suggesting a moderate correlation between the two variables. This might suggest some link between the phrase length and WPM. Yet, I expect to see this relationship plateau at some limit.

None of the participants in my study had used a re-speaking feature before. Although I did receive some positive feedback, some participants also mentioned that they would prefer re-speaking the entire phrase, rather than checking the recognition and locating the erroneous part.

From my qualitative results, I noticed that participants generally exhibited a relaxed attitude toward the errors they encountered. When I asked for a comparison between my system and other voice-to-text systems they had used, participants often brought up scenarios where other systems make errors. The general consensus among participants was along the lines of *"Although your system sometimes makes errors, other systems also make errors—even more errors. So your system isn't that bad."* This suggests that participants are more used to errors in speech transcription systems.

From the interview results, I also found that users noticed some imperfections in the re-speaking feature, especially surrounding issues with the matching algorithm (even though this occurred only rarely). Indeed, a re-speaking feature can be prone to alignment issues [53], which could then be confusing for users as they have no indication of how much text they need to re-speak [17].

Overall, my results suggest that the insertion of errors in speech transcription might have some effect on the entry efficiency and perceived workload. However, in contrast to my expectations, this effect seems to be small. My results suggest that, compared to typing-based text entry tasks, users seem to have a higher tolerance to errors in speech transcription tasks.

In real-world uses of speech recognition such as Google Home devices which do not have a display, speech-to-text users usually do not or cannot look at the output, making it more challenging to correct the text. To replace the need for a display screen as used in my study, system designers might consider integrating a trigger word such as "sorry" to activate a re-speaking attempt. Then, after users say the trigger word, a correction attempt would be automatically activated and users can then proceed with the correction without having to look at the recognized results.

In the future, I plan to re-evaluate my system using other error correction methods, such as direct voice commands. I also plan to include a third attempt for error correction, as suggested by some participants in the interviews.

Another potential future direction of my work is to investigate speech-to-text behaviours in different contexts. Recent research by [2] studied users' typing performance in the wild, and they found that different contextual factors can have an impact on users typing speed and error rate. Given that my studies were done in a controlled lab setting, I would thus also like to investigate how different contextual factors may affect users' speech-to-text performance.

Chapter 7

Conclusion

Through two user studies on speech-based transcription, I explored the effect of errors on entry efficiency and subjective experience, using both quantitative and qualitative measures. I evaluated the efficiency of a re-speaking feature for error correction. I found that users are more likely to experience higher frustration and mental workload as the rate of errors increased. In addition, based on my findings for the second verification time (VT2), I also identified that the re-speaking feature is likely not sufficient for error recovery, especially when a high number of errors occurs in the system. While the act of inserting artificial errors into recognition results did not exert a significant direct impact on participants' performance in my work, it nevertheless affected their efficacy in detecting inaccuracies. This outcome may be indicative of a reduction in certain cognitive faculties or emotional states, potentially arising from the decreased reliability of the system as a whole. This is again a topic for future work. Overall, my findings suggest users have a fairly high tolerance for errors in speech transcription. The main outcome of this research is thus potentially good news for speech interface designers and programmers, as a limited amount of errors in the systems might not affect the user experience in a substantial way.

Bibliography

[1] F. Niyi Akinnaso. On the differences between spoken and written language. *Language and Speech*, 25(2):97–125, 1982.

[2] Elgin Akpinar, Yeliz Yeşilada, and Pinar Karagöz. Effect of context on smartphone users' typing performance in the wild. *ACM Trans. Comput.-Hum. Interact.*, 30(3), jun 2023.

[3] Ohoud Alharbi, Ahmed Sabbir Arif, Wolfgang Stuerzlinger, Mark D. Dunlop, and Andreas Komninos. Wisetype: A tablet keyboard with color-coded visualization and various editing options for error correction. In *Graphics Interface*, GI '19, May 2019.

[4] Ohoud Alharbi and Wolfgang Stuerzlinger. Auto-cucumber: The impact of autocorrection failures on users' frustration. In *Graphics Interface*, GI '22, pages 148–160, May 2022.

[5] Ohoud Alharbi, Wolfgang Stuerzlinger, and Felix Putze. The effects of predictive features of mobile keyboards on text entry speed and errors. *Proceedings of the ACM Human-Computer Interaction*, 4(ISS), Nov 2020. ISS '20.

[6] Ahmed Ali and Steve Renals. Word error rate estimation for speech recognition: e-WER. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, ACL '18 (Volume 2: Short Papers), Jul 2018.

[7] Shiri Azenkot and Nicole B. Lee. Exploring the use of speech input by blind people on mobile devices. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '13, Oct 2013.

[8] S. Bagnara and A. Rizzo. A methodology for the analysis of error processes in human-computer interaction. In *Proceedings of the third international conference on human-computer interaction, Vol. 1 on Work with computers: organizational, management, stress and health aspects*, HCII '89, Sep 1989.

[9] Timo Baumann, Casey Kennington, Julian Hough, and David Schlangen. Recognising conversational speech: What an incremental ASR should do for a dialogue system and how to get there. In *Dialogues with Social Robots: Enablements, Analyses, and Evaluation*, pages 421–432. Springer, Dec 2017.

[10] Gamal Bohouta and Veton Këpuska. Comparing speech recognition systems (microsoft API, google API and CMU sphinx). *Int. Journal of Engineering Research and Application*, 07(03):20–24, Mar 2017.

[11] Stuart K. Card, Thomas P. Moran, and Allen Newell. Computer text-editing: An information-processing analysis of a routine cognitive skill. *Cognitive Psychology*, 12(1):32–74, Jan 1980.

[12] Christine Cuskley, Francesca Colaiori, Claudio Castellano, Vittorio Loreto, Martina Pugliese, and Francesca Tria. The adoption of linguistic rules in native and non-native speakers: Evidence from a wug task. *Journal of Memory and Language*, 84:205–223, Oct 2015.

[13] G. E. Dahl, Dong Yu, Li Deng, and A. Acero. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1):30–42, Jan 2012.

[14] Mark Davies. English corpora. `https://www.english-corpora.org`, 2023. Accessed: 2023-01-12.

[15] Adele Diamond. Executive functions. *Annual review of psychology*, 64, 09 2012.

[16] John W Du Bois. Santa barbara corpus of spoken american english | department of linguistics - UC santa barbara. `https://www.linguistics.ucsb.edu/research/santa-barbara-corpus`, 2022. Accessed: 2022-6-21.

[17] Jiayue Fan, Chenning Xu, Chun Yu, and Yuanchun Shi. Just speak it: Minimize cognitive load for eyes-free text editing with a smart voice assistant. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, Oct 2021.

[18] Margaret Foley, Géry Casiez, and Daniel Vogel. Comparing smartphone speech recognition and touchscreen typing for composition and transcription. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–11, Apr 2020.

[19] Debjyoti Ghosh, Pin Sym Foong, Shengdong Zhao, Di Chen, and Morten Fjeld. EDITalk: Towards designing eyes-free interactions for mobile word processing. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–10, Apr 2018.

[20] Debjyoti Ghosh, Pin Sym Foong, Shengdong Zhao, Can Liu, Nuwan Janaka, and Vinitha Erusu. EYEditor: Towards on-the-go heads-up text editing using voice and manual input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, 2022-07-13.

[21] Debjyoti Ghosh, Can Liu, Shengdong Zhao, and Kotaro Hara. Commanding and redictation: Developing eyes-free voice-based interaction for editing dictated text. *ACM Trans. Comput.-Hum. Interact.*, 27(4), aug 2020.

[22] Google. Speech-to-text: Automatic speech recognition. `https://cloud.google.com/speech-to-text`, 2021. Accessed: 2021-12-15.

[23] Christine A Halverson, Daniel B Horn, Clare-Marie Karat, and John Karat. The beauty of errors: Patterns of error correction in desktop speech systems. In *IFIP TC13 International Conference on Human-Computer Interaction*, INTERACT '99, Jan 1999.

[24] Morten Hertzum and Kasper Hornbæk. Frustration: Still a common user experience. *ACM Trans. Comput.-Hum. Interact.*, 30(3), jun 2023.

[25] B. Juang and Lawrence Rabiner. Automatic speech recognition - a brief history of the technology development. *Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara*, 01(67), 2005.

[26] Clare-Marie Karat, Christine Halverson, Daniel Horn, and John Karat. Patterns of entry and correction in large vocabulary continuous speech recognition systems. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 568–575, May 1999.

[27] Veton Kepuska. Wake-up-word speech recognition. In *Speech Technologies*, chapter 12. IntechOpen, 2011.

[28] Takashi Kimura, Takashi Nose, Shinji Hirooka, Yuya Chiba, and Akinori Ito. Comparison of speech recognition performance between kaldi and google cloud speech API. In *Recent Advances in Intelligent Information Hiding and Multimedia Signal Processing*, SIST '19, pages 109–115, Jan 2019.

[29] H.H. Koester and S.P. Levine. Modeling the speed of text entry with a word prediction interface. *IEEE Transactions on Rehabilitation Engineering*, 2(3):177–187, Sep 1994.

[30] Can Liu, Siying Hu, Li Feng, and Mingming Fan. Typist experiment: An investigation of human-to-human dictation via role-play to inform voice-based text authoring. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), Nov 2022.

[31] Donald G MacKay. Awareness and error detection: New theories and research paradigms. *Consciousness and Cognition*, 1(3):199–225, 1992.

[32] I. Scott MacKenzie and R. William Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 754–755. Association for Computing Machinery, Dec 2003.

[33] Arthur E. McNair and Alex Waibel. Improving recognizer acceptance through robust, natural speech repair. In *Proc. 3rd International Conference on Spoken Language Processing (ICSLP 1994)*, pages 1299–1302, 1994.

[34] Fabrizio Morbini, Kartik Audhkhasi, Kenji Sagae, Ron Artstein, Doğan Can, Panayiotis Georgiou, Shri Narayanan, Anton Leuski, and David Traum. Which ASR should i choose for my dialogue system? In *Proceedings of the SIGDIAL 2013 Conference*, SIGDIAL '13, pages 394–403, Aug 2013.

[35] Christine Murad, Cosmin Munteanu, and Wolfgang Stuerzlinger. Effects of wer on asr correction interfaces for mobile text entry. Poster, Oct 2019.

[36] Hay Mar Naing and Win Pa. Automatic speech recognition on spontaneous interview speech. In *Sixteenth International Conferences on Computer Applications*, ICCA' 18, Mar 2018.

[37] Sharon Oviatt, Antonella DeAngeli, and Karen Kuhn. Integration and synchronization of input modes during multimodal human-computer interaction. In *Referring Phenomena in a Multimedia Context and their Computational Treatment*, ReferringPhenomena '97, pages 1–13, Jul 1997.

[38] Philip Quinn and Shumin Zhai. A cost-benefit study of text entry suggestion interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, pages 83–88. ACM, May 2016.

[39] Brenda Rapp, Simon Fischer-Baum, and Michele Miozzo. Modality and morphology: What we write may not be what we say. *Psychological Science*, 26(6):892–902, 2015.

[40] James Reason and Dieter Zapf. Introduction: Human errors and error handling. *Applied Psychology: An International Review*, 43(4):427–432, 1990.

[41] Shyam Reyal, Shumin Zhai, and Per Ola Kristensson. Performance and user experience of touchscreen and gesture keyboards in a lab setting and in the wild. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 679–688. ACM, Apr 2015.

[42] Paul Rozin and Edward B. Royzman. Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, 5(4):296–320, Nov 2001.

[43] Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay. Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–23, 2018.

[44] Karpagavalli S and Evania Chandra. A review on automatic speech recognition architecture and approaches. *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 9:393–404, Apr 2016.

[45] Haşim Sak, Andrew Senior, and Françoise Beaufays. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv:1402.1128 [cs, stat]*, Feb 2014.

[46] Korok Sengupta, Sabin Bhattarai, Sayan Sarcar, I. Scott MacKenzie, and Steffen Staab. Leveraging error correction in voice-based text entry by talk-and-gaze. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–11. ACM, Apr 2020.

[47] Cristina Simões-Franklin, Robert Hester, Marina Shpaner, John J. Foxe, and Hugh Garavan. Executive function and error detection: The effect of motivation on cingulate and ventral striatum activity. *Human Brain Mapping*, 31(3):458–469, 2010.

[48] Linda Skitka, Kathleen Mosier, and Mark Burdick. Accountability and automation bias. *International Journal of Human-Computer Studies*, 52:701–717, Apr 2000.

[49] R. William Soukoreff and I. Scott MacKenzie. Measuring errors in text entry tasks: an application of the levenshtein string distance statistic. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, pages 319–320, Mar 2001.

[50] Bernhard Suhm and Alex Waibel. Exploiting repair context in interactive error recovery. In *Proc. 5th European Conference on Speech Communication and Technology (Eurospeech 1997)*, ISCA '97, pages 1659–1662, 1997.

[51] Steve Tauroza and Desmond Allison. Speech rates in British English. *Applied Linguistics*, 11(1):90–105, Mar 1990.

[52] Keith Vertanen, Dylan Gaines, Crystal Fletcher, Alex M. Stanage, Robbie Watling, and Per Ola Kristensson. VelociWatch: Designing and evaluating a virtual keyboard for the input of challenging text. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, May 2019.

[53] Keith Vertanen and Per Ola Kristensson. Automatic selection of recognition errors by respeaking the intended text. In *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, ASRU '09, pages 130–135. IEEE, Dec 2009.

[54] Keith Vertanen and Per Ola Kristensson. A versatile dataset for text entry evaluations based on genuine mobile emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, page 295–298. Association for Computing Machinery, 2011.

[55] Keith Vertanen and Per Ola Kristensson. Complementing text entry evaluations with a composition task. *ACM Trans. Comput.-Hum. Interact.*, 21(2), Feb 2014.

[56] Yuntao Wang, Ao Yu, Xin Yi, Yuanwei Zhang, Ishan Chatterjee, Shwetak Patel, and Yuanchun Shi. Facilitating text entry on smartphones with QWERTY keyboard for users with parkinson's disease. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, pages 1–12. ACM, May 2021.

[57] Dieter Zapf and James T. Reason. Introduction: Human errors and error handling. *Applied Psychology*, 43(4):427–432, 1994.

Appendix A

Complete Phrase Set Used In Study 2

Seems like he could've just reached over and took it.

You must get him back to Earth as soon as possible.

I had to slow down the videotape to frame by frame to see how it happened.

You know, I don't think I would.

We've done everything we needed to do.

Everybody thinks they've got a problem.

Finally he got to the other side, and the boy told him who he was.

Honestly, you don't know what you're talking about.

I think I was concerned about you getting there and back.

I think it's a very serious message, especially at a time of economic downturn.

Obviously, you haven't learned your lesson until you get punished for what you did.

Nice little profit center, isn't it?

Like it wasn't completely obvious the way you ripped apart his documentary.

I got a couple more things for you to do.

I mean, it's not the drilling so much.

He could never do what needed to be done.

Like just watching you do the things that you love when you don't know that I'm looking.

You know, I wish I could have had more, but you and your sisters didn't even get that, and now she's back.

What are you talkin about? Listen to me.

What do you want for Christmas?

Seriously, if people would just leave me in peace and let me do my thing, we would all get along just fine.

Well, I was just calling a cab, father.

When she comes back here tonight, if she does, I'm going to make damn sure that's just

another empty promise.

Disembodied heads in the ice chest don't exactly make for five star ratings.

You know? Luis misses her so much already.

The cranes kept your family from finding work in harmony just to get by after your father disappeared.

I'm telling you, they are on a scoring rampage now?

Can't wait to see you, tell you face-to-face how proud I am.

God forbid you should actually spare one drop, one actual drop of love from the children.

No, I can't tell you on the phone.

Standard roadside checks within 50 miles of the border.

Sweetie, I told you not to play in here until we clean it up.

This costume happens to be a protest statement.

I get the impression it wouldn't be that hard to hide something in Connecticut.

So that means he can't be seen with a sophomore? No, it means he doesn't even see sophomores.

You're trying to think of a way out of this, but I can't take any more crap tonight.

Sam must have had a really bad day.

We're talking about the United States Supreme Court.

Majority of us have been here a long time.

How is she? She's not going to lose her mind.

Hey, I got to ask you something.

You guys haven't bought your tree yet? There are only two days till Christmas.

Austin's company is moving in on High style.

Doesn't it automatically go off at the end?

Nothing as big as this ever runs smoothly, anything worthwhile is worth fighting for.

We're going to make the most patriotic rice crispies anyone has ever seen.

How does it feel? How does it to hurt you?

You have hardly seen hide nor hair of him for months.

Move the arm and the shoulder outward till you hear it click in.

I feel like I should get you another sweater.

You're supposed to be a neat freak.

It seems we're having a winner of the week in this race.

Cathy calls me up, and she wants Johnathan's address.

It's really built an amazing society in its short history.

I feel literally physically lighter.

Good luck finding an indispensable face.

What are you still doing here? We heard you were fired.

If she had to go shopping or something maybe you could go with her.

But no matter how much you love curry, you had too much of it.

They shouldn't be all dry and cracked and crunchy around the edges like that.

If the blend falls short, you will probably be the one that could be.

And the place has incredible potential and somebody is gonna get a chance to make it a wonderful inn.

I'm just confused why we're not enough for you.

I don't have the time to give you a lot of statistics.

He's going to search through everything Arianna's ever touched until he finds that camera.

The works festival event with over 45 exhibits at over 25 sites downtown.

Okay, then I will keep doing that.

Yes, the quality of the knights, the noble heart, boldness over justice, integrity.

Speed limit in a school zone is 30, not 50.

That's five bucks a word for the inscription and ten for the signature.

A safe, strong, secure Bridget.

Why do you wear them? The sun has been down for an hour.

Damn it, Mike, we can barely afford the mortgage.

The book mirrored her own life in more ways than she ever imagined.

You got to leave him a plate full of insulin.

In both instances, an item was taken to be returned to its rightful owner.

You hear me? I mean, those were your exact words.

I came up with an alternative idea.

I mean, he was right there in the front of the pack most of the time with that microphone.

Get out the gold stars, we have come up with a plan.

This is a difficult section of the track, very bumpy and high speed.

Mom has this beautiful dress where the straps are made of diamonds.

The president has got a whole series of measures that he's going to unveil tomorrow.

We can basically reconcile the science with what's in Genesis.

Your mother phoned and said you were coming to visit.

What do you say? I say that it's important that whoever reps you cares about you, right?

I thought you had to stay to get ready for your play.

At age 62, I just purchased the first new vehicle in my life, a Prius.

One little pill, that's all it would take, and you would feel so much better.

Anything you say can and will be used against you in a court of law.

Can't be responsible for him all of your life.

I mean, how is that anything more than just a strange coincidence?

Waiting for me? What do you mean? Why would you be waiting for me?

No objection to the unintelligible authorities.

Your father can be very generous, but he can also be very vindictive at times.

He's become a trusted friend, you know? Someone I can rely on.

He seemed so committed to the job.

Listen, I know I made fun of you in the past for riding the bus, but I have to admit I was wrong.

I'm just the caretaker, I guess you could say.

So I guess it wouldn't matter if I just jump off the wagon completely, since you already think I'm making pit stops.

What do you mean?

I didn't really expect him to linger.

All the man wanted to do was build some eco friendly houses in Malibu.

Got to take a quick break right now.

If you can poke a hole in your busy schedule, then I could buy you lunch.

In the meantime, I got a club here that needs a new image, and I got a singer with a golden voice who's dying to use it.

I didn't mean anything by that.

Time for you to get to know and trust the crew, and time for them to understand how you want things done.

I just want you to give a little thought to this moat.

Instead, I'm sitting here wondering if my music career is over before it even started.

I did all the calculations myself.

I wanna rewind it and hear that back.

You know, he just didn't want you to sue him over the car accident.

I feel those reports would be useful to me.

You had your chance with Theresa and you blew it.

It was broken,we were gonna take it out and send it back to the factory.

I'm sure things will get better for you.

My husband is a huge hockey nut.

Less than one in three say this is important.

But it was nothing like the Beatles, even the Stones, you know.

I'm sorry, where's the harassment? She's part of an ongoing investigation.

We all have a role to play at Forrester.

It's hard for me to understand, but it happens, and that is what the Bible teaches.

He didn't see that way and ultimately I said, look, Mr.

I wonder if we should wake her up?

I remembered him from a year ago.

So that means he can't be seen with a sophomore?

Here's the police station, fire station, you must think we're totally paranoid.

But if you have doubts and opinions, I really wish that you would keep them to yourself? Please.

I don't want to ever go to bed upset.

If that is not acceptable, I will be willing to quit, teacher.

I want to go off someplace with a bottle of vodka and blot out the entire world.

I will remember to set the timer.

Talk to her about being a lawyer, okay? I can do that.

You changed it from a place of tragedy to a place of hope.

Have you read through your statement recently?

We still have 78 more channels to go.

I want you to make yourselves comfortable.

I was just waiting for the right man to come and sweep me right off my feet.

I thought I was going nuts.You know, I read that it takes all of the nutrition out of our food.

She can't be late to start and I'm feeling guilty for not thinking of it sooner.

She managed to pick up the pieces of her life, very bravely, I might add.

You're not going to be able to do that by nickel and diming me in the budget.

But he's just really really really strange.

If you can hang out for a bit longer I can say a proper goodbye.

The Mouse King is too powerful to take on alone.

To love, honor, and cherish forsaking all others till death do us part.

Should we have an attorney to advise us.

Let's just take a quick break.

From what I know of Katie, there's no reason that she would have done something like that.

He doesn't realize that.

As his friend, I'm asking you to stay.

We will start with you, Alexander.

Maybe she just thought she would keep it quiet, make it a little more simple.

You're nervous to be at a new school because you think no one can relate.

You can call me when it's hot, fast and fresh.

We've talked to other people like Julia, people who don't normally come out in protest for various events.

Look, I just want to say I'm sorry for all that stuff.

I couldn't do that Anna, otherwise I have no money left.

You were saying that you're leaving to nurture your inner child.

Am I doing that right so far?

I won't be staying long at all.

I want you to see that I love you more than anything in the world.

I guess we have to put those in the oven, if we're gonna eat them.

I started my own first band when I was 45.

In the wake of these and other reports, with a strong partisan tilt to the assumptions.

Too bad we're going to miss it, because we will be on a plane off to our wedding to make it official.

I still think you're being ridiculous about the whole thing.

You got me a cola drink? And a lemon lime? Well, this is too much?

I'm having like this really great overlay, you know, which is the ferry.

I never saw him before in my life, right, Ed? I will see you guys later.

Soften you up in a whole new cell with loud music, deprive you of sleep.

Our selections tonight are chicken piccata or seafood gumbo.

I think we're out of our comfort zone here, you know? This isn't working out the way we imagined it would.

There's no way I'm gonna let her touch my hair a second time.

What's left has been assigned to me.

No, no, just fine, take your word for it.

But he needs to stay off of it for six to eight weeks.

Maybe I better just take you along for the ride.

Because then I could wake up, and I could be myself.

You can't really tell when they blush, it's very unusual.

He's striking up a connection with the Mexicans.

I can't even believe that I let them talk me out of it.

We agreed to meet with Paul but, quite frankly, the decision has already been made.

But now that you're on your own, I get to deal with both of you.

But you were a little more right.

So everyone kind of chipped in for the surprise for her.

I like to think of myself as a problem solver.

Although you may want to check in on her.

It sounds so fancy!

An urgent appeal on behalf of millions facing an uncertain future.

How's he doing? Stable for the moment.

Why don't we get more of that rewards?

He's the reason the warrant came so easy.

Geez, I don't want to hurt anybody.

They were paid for with your credit card.

When we can't accept that we've died, we create the illusion of life to hold on to.

And what about the annulment papers he's always offering to sign?

They came up through the soft dirt.

They have about thirteen hundred employees still in the Houston area.

Lisa has done a lot of reporting on this.

Lucky! I wish I grew up in a place like this.

I never said anything about the wheels staying on.

But right now, you need to stop thinking about other people and their feelings, and you need to focus on you.

What extraodinary things you do say, Lucy?

It's not gonna work that way.

I was so glad that he opened up on this disclosure bit.

Well, you're gonna have to come up with a better argument than that if you go downstairs.

You ask someone why they're interested in electronics,

I hear that you have done an excellent job with the Kline project.

But it was extraordinary to find somebody like that as an eyewitness to such an event.

The trouble is, Jessica doesn't like my kind of fun.

No trouble, but I warn you, it's quite a walk.

I guess our children must be worried about us.

You attach it to the input on your television and it basically blocks news transmissions.

Why don't you take these off of me and see how that works out for you?

I am sorry, but it's just not my style.

He wore a new bow tie and a different bag.

Look, you know what? Look, we can talk in the morning.

She walks in, it's like a tornado.

I'm gonna wait till like early in the morning.

I don't mean to disturb your hard work.

I mean, this sounds like something you see in a movie.

I use over-the-counter remedies.

Chief, look? No one knows what caused it.

Unfortunately, they were corrupted during the melting.

Plus, there's nothing on my screen about it.

Did you give candy to the kids?

They were probably very nervous when they came in just for jury duty.

I dropped her off at the bookkeeper.

They said, "We have received your question about U."

Just wondering if you're there, if you cooled down a little bit.

And you thought she would done it, didn't you? When you first arrived on the scene.

She wants nothing but the best for herself, for her family, for all of us.

It's about one o'clock.

We almost got a divorce until we realized how much we love each other.

I thought you were going to get out at the square.

He put out this 25 to 30-year record, and people responded to it.

Look, maybe you guys should be out there looking for Janice right now instead, you know?

Correct me if I'm wrong, didn't they invent something called an elevator about 100 years ago?

There have been symbolic resolutions supporting NATO.

Let's just listen in, and see what her remarks are on.

I trust you recognize, we cannot accommodate anyone else.

But then you told me you wanted me to be comfortable, so I wore that instead.

If anything happens to this volume between now and Monday, I'm holding you responsible.

One doesn't always stay young and perfect.

I don't want to ruin the party, but I wanna say something.

His brain is deteriorating at a speed you will find shocking.

Well I was gonna make a sandwich with the leftovers.

She just taught me a lot of things that a mother would teach a little girl.

People are going to be shocked to see the numbers.

That turned out to be more difficult than either of them had imagined.

The profession you have dishonored is ashamed of you.

Paper said your husband was a lawyer in New Hampshire.

I know that there's more important things in life, but then again, maybe there aren't.

If I don't call in by nine o'clock, I will be glad to do that for a fee.

It's not like she doesn't know something's up.

Politics is tied into the whole concept of power.

Here, you want the tickets? Cory, he paid for those tickets with his own money.

I don't think it will come as any great surprise to you that I stopped reading newspapers.

They made plans to meet after work.

Paul is a volunteer on the orphanage project.

I was thinking the very same thing.

Like the sun and the moon, only one of them can be in ascendance at any given time.

I'm really close with Kathy, even more so, because I know what she went through.

They were getting married, Will was finally going to get his parents together.

Are you gonna ask Kevin to be your best man before dawn?

Kev, where's your new girlfriend?

If I play my cards right over the next few years, we're not going to have to worry about money ever again.

I am checking both of your blood for the presence of the organism.

Maybe I'm just eager to get back to where I was.

I just want to enjoy that, you know?

The question is the auto industry is going to have to innovate here a little bit.

You still haven't touched your food.

Trying to get me drunk? I tried to pretend it didn't happen.

She's a very talented designer.

He can't access this computer, it's not hooked up to a phone line.

I think it would be cool if we put them together to create one bug.

We're going to open this natural food supermarket, one of the first natural food supermarkets anywhere in the world.

He's asked that we evaluate the scientific and cultural value of each mission from now on.

By all means, finish mine.

I'm just worried about how he's going to deal with this.

You know, just like a cat, regular cat.

I left her a message at home and on her cell, but she hasn't called me back.

Oh? And fixings for a few meals.

I wrote it down on the back of an envelope for you.

And the fact is the military just overturned the civilian government.

I didn't mean to startle you, but you know, I was worried about you last night.

I saw a picture of this gal in the newspaper at some society wing ding.

That producer and you seemed pretty intense.

I am sorry if you have all gotten the impression that I like Matthew more than the rest of you.

I love you more than anything in the world and I will always be with you, OK? Remember that.

Listen, we haven't even been here that long.

But with a care, a man just might survive it.

I was like Jeff, I called him and I said I'm just not meant to be a mom.

Yes, once Theresa is out of the way, she won't be looking over Chris's shoulders.

Well, just try to be a little more careful next time, okay, kids?

More than 20 million Americans have received coverage under the law.

My school principal attached these devices to all the televisions in the building.

How do you think they are as two characters together?

I will do everything I can to help out.

I keep sending messages home with their folks, saying, please, you know,in Spanish.

As the guy who tells her the truth, no strings attached.

Matter of fact, I dream about you every night.

He's a tremendous genius, he's a great intellect.

Everybody wanted to play with Mary.

I would have thought the same thing.

But also completely undetectable by any of today's standard drug tests.

We have it in four colors and it's just very charming and simple and perfect for the summer.

There are huge long waiting lists for classes.

Well, our local paper is not so bad.

Now I can't imagine my life with anyone else or in another way.

I want you to access the transformation program and, when you're ready, input the symbol.

Tell you what, we will take care of this.

What's your reaction to this ad? Well, it is horrible.

This peculiar beast, which actually has the right incentives, you just seek the truth.

I have been on decaf for years.

That is why Project Blue Demon is of the utmost importance to our government.

With the end of all of the negative things, there's a new beginning.

You might want to try the front porch.

We're about to take a trip around the world.

You want to tell him about the dogs or should I?

I mean, this guy is just beggin to be made an example of.

Reverse the situation, and you didn't hesitate.

Pope earned his doctorate in philosophy from Seattle Pacific.

You know, I have a feeling this one's going to be even better than the last.

Perhaps you will find what you're looking for here, too, Kendall.

Go? Think of the publicity we will get.

I kept calling the office and saying do you have this form, and they would tell me no.

In the end, she sewed that little bird with it.

How old are you? You look about 15.

Only thing I'm racing now is the expiration date on our milk.

We will see the president raising money.

When I'm up there as Hannah Montana, I get nervous, too.

Crucially, describe how a biological system could self organise.

I'm gonna keep my mouth shut, and I'm going to talk to myself.

I think that it will be interesting when you hear from Mr.

You know, you're very easy to talk to, and that goes a long way with women.

We will talk more after a break.

We have to make a decision here.

Mike, I see after our dialogue that you're a very idealistic person.

Christmas Eve, that's five weeks from today.

There are plenty of opinions about this.

It went on for three or four months.

We were kicked out of the house.

We didn't want to alarm you.

Never forget, the blood of the courageous James clan flows through your veins.

Why don't you join them?

You're gonna give them to me, or I'm gonna get them from you.

This is what I painted while I was on amphetamines.

He spends three months every year organizing the festival, dealing with every detail.

It's just not safe for a woman by herself.

What's a park ranger vehicle doing here?

So we were there for many, many years, and they killed more in the months following, as you know.

People probably won't even care to look it up.

This isn't the first crisis I've been in.

I wish you could reconsider my offer.

We don't have to talk about it at lunch it you don't want.

But, as office manager, she handled everything delivered to the practice.

They live on through the deeds associated with them.

The only time you come in here and eat pie is when you're depressed.

Apparently it's not working so I'm gonna switch her to a stronger antibiotic.

It all happened so fast, I must have totally misunderstood.

I was supposed to go out, but it was snowing.

Know where it is? I will find it.

So we need to get with it and we need to start pronouncing it right.

Melanie needed me to stop by, and let everybody know that we're going out of town.

If word of them ever got out, no one would ever settle here.

Really? I didn't hear about it.

Once it's off the newsstands, it's in the recycle bin like it never even happened.

Whatever is easier for you guys.

It could have been that she herself was immature.

Do you want to play hero? It's time to live up to your word.

Ok, what kind of security are we talking about here?

Depending on the dosage, it can bring on a stroke, or massive cardiac arrest.

I think it's very important for us to talk about reality, as well.

I haven't read the book so I don't know.

Guys, be careful when you're walking on the rocks.

They didn't announce the winner, they just started announcing the cast.

I won't be going to the football game.

I don't know who comes up with these names.

I mean, it's good that she's helping Yolanda.

You don't have to give me an answer right now.

It's inside, but I don't want to call the police, because I don't trust him.

I've played in the NFL, remember? I've handled groupies, crazy fans, you name it.

You know, that the cards were essentially dealt to us before we even reach the table.

Are you saying all it took to kill this man was the equivalent of a dozen or so grains of salt?

I do not want anyone else to die in the Balkans.

I don't think it's a good idea for you to go up there in the winter.

I know you have probably gone on with your life and that's fine.

I'm just finishing up in the kitchen.

Look, I hate to bring this up cause I'm really close to the guy.

His decision was not overwhelmingly surprising, it was overwhelmingly disappointing, though.

You know, it's not as hard as you might think and you can do it.

Guess the history tapes weren't so far off after all.

And now that you're here, I finally have the one person who can make my plan a success.

What are you doing? Sam, what are you doing?

I think he's more than capable of being able to count backwards five or six months.

My doctor thought it was best I stay out of touch for a while.

I'm watching Jen, and you're going to be here in five minutes.

If it's ok, I will just wait because I could really use a friend right now.

But then maybe she needs to take the whole family and leave and never come back.

This is a real foodie and she gets to meet all these cool people who have signed your plates.

This weighs a ton, and, you know, I didn't see any bellboys near here.

Okay, I can see you are excited but what about work?

Well, this is like a chess game I have to anticipate ten steps ahead of him.

I've never in my life done anything where I don't do it exceptionally well.

Well, it must be coming from next door.

And I could give you half a dozen other examples.

Don't worry your pretty little head about it now.

What's up with that dude from the bank?

You need every edge you can get.

Having a good day? Thanks for the coffee.

How do you do, Mr Allen? Number 41.

What if that ambulance was carrying somebody to the hospital?

I want our students to be prepared.

Why don't you step off?

It's just that the last belly dancer.

How about an extra three days? How would that sound?

So it's no surprise that she sat down and wrote Dorothy a letter.

And if they stay here in harmony, they're going to keep crossing it over and over again.

However, we must go into the matter a little more carefully, I'm afraid.

You know where Sarah and Jerry live?

It did turn out okay after all.

She would come to dinner rather frequently, and we always had a lovely evening.

I feel like you guys are so busy.

The car engine has to be replaced.

Let's all get together again!

Once we had those, we were able to grow much faster.

I haven't used the truck in a couple of days.

You better wake up and start being honest with yourself before it's too late.

Suspecting that poor sweet man of stealing my jewellery.

Stop looking at me like that.

Meanwhile, the weather system continues to bring destruction throughout the region.

I think we're having some technical difficulties.

One of those Olympic moments to remember.

That planet's been an endless source of grief for me.

But you're the future that I've always wanted, and I'm not going to let anything jeopardize that.

Are you freaking kidding me? Get upstairs.

I could use some professional advice.

I would say, you always look amazing.

Fatty food, ugly decadence, and rigged games.

Because I think that it would be most expedient.

Maybe I could go and get back before he wakes up.

We're in a downward spiral right now, Luke.

Come on, Harry? Look at me? Nine, ten, eleven, twelve.

They're getting on and off the metal structure.

Their monthly payment of principle interest taxes and insurance would be seven hundred.

The publications committee needs to be more involved than just making decisions on what we publish.

Two pairs of cotton pajamas.

I probably don't want to know the answer anyway.

Let me, first of all, talk to you about the concept of a timeline.

They will try to distract you,stay focused.

Where did that come from?

There's been a small miscommunication and they've been moved out into an enclosed high security lagoon.

You had it in you the whole time.

God, you could destroy the vortex completely.

I wish you had gotten in touch.

You know, that's when you coming to an end.

Thought maybe you would like to take him to school? Yeah.

All right, I will have to get it detailed, but what matters is that you're all right.

The other guy was shorter, darker hair.

It's possible we are seeing the beginning of that now!

What's that mean? You're next in line.

Ok, we got out of the stairwell, we went into the corridor.

Bro, are you making house calls now?

What are the risks versus the rewards?

He had surgery for cervical cancer in 2005, but without the recommended follow-up.

I didn't think anybody would believe me because he is the one that had all the money.

Need a lift? Where are you headed?

You'd think with a scholarship, to Cornell wouldn't be a no brainer, but no.

But a hell of a lot of feelings died along the way, and the only ones I had left, I didn't know how to use.

Well, Jesus, you have been sober for 15 years.

Get mom's phone from the school and track the damn thing.

I thought that we could go to the park and have a picnic lunch like we used to.

The meaning of these events was astounding.

A bit of wind up material on the opposition.

We have to sort of find interesting aspects of our public figures that draw us to them.

This was going to be his first Fourth of July as a resident of the United States.

I wanna wish everyone a happy holiday.

You should become sorry for yourself but then the song should pick you back up.

Where he says, in the face of despair, we can still hope.

OK, are you about done in there? Hold on.

I was just gonna use a tin can and put rocks in the bottom.

Sweetheart, I'm sorry that you have all these feelings right now.

He doesn't have a long-term fix.

You're not lookin at what I'm lookin at.

I will just have the fruit salad.

I thought I would get started sorting through the materials.

See? Frank didn't like that.

Zoe, come on, it's freezing out here.

So if you feel other to help them alleviate it.

Billy, the doctor told you he wants you to avoid a stressful environment.

I mean, I thought it was over between you and Maria.

The entire store of human knowledge now doubles every five years.

Do you wanna come with me? What are you doing?

Sensitivity to sunlight is a side effect?

If you're gonna poke the hornet's nest, poke the drone.

Always add everything up on the calculator first, just to make sure the figures match.

I mean, you really ought to be thanking her.

Sabra is a Hebrew word.

I will say that I think they work.

He lives a dangerous life and suffers for it and so do the people around him.

I'm going up to the mountains for a couple of days.

We will launch a class 5 probe just before we reach the border.

The songs they were playing on the radio were the exact same songs they were singing at

party.

Because it wasn't that easy for him to get around.

And 100 years from now, who knows what's possible?

Drop it? Put your hands on the back of your head.

I'm not interested in doing that much work.

And with the whole town needing to be fixed, take a look at your engine for you.

It was just really the wrong move there.

I had plans to go to Amy's game tonight.

You tend not to change your mind about how you feel about things.

Let's rewind, and you can ask me the secrets question again, okay?

Sometimes true stories seem hard to believe.

Secrets don't always stay buried.

You cut us a deal and I'm going to make you a very wealthy, powerful man.

It wouldn't surprise me if he's nearby, watching and listening as we speak.

Come on, Molly, your seasonings could be in every grocery store across the country.

Hopefully it's good news.

Why? You said somebody is got to take his place.

I thought it was just so wacky, a little bit of it.

When our viewers see you on the screen showing them how to be a hero and not a victim, they'll feel empowered.

This will tear his heart out, then again, maybe it's not what it looks like.

He, we hope, will be joining us tonight from Louisiana.

Waiting lists for these types of transplants can last from days to months on end.

This epilogue is really fresh and riveting.

I said that there's something familiar about him, so I'm going to go talk to him and find out what it is.

At the gate, many were waiting when the weekend ended.

But then later, Kendall had to tell him about Zach.

Look, you have come at the wrong end.

You weren't supposed to go inside.

I'm playing with fairly expensive money.

So I know you think I'm crazy, but I really need the money so I'm going back in there.

Thank you! That's very nice of you.

I want you to remember me the way that I was before, because that's who I really am.

Come on now, you must get up, coz breakfast is ready.

I have to go back to the tree house and get Veronica, and a few items we might need.

Some kind of alien information has been downloaded into our computer core.

The bill we had in front of us was tough.

There was a lot of disappointment going around that day.

That doesn't mean that you can pretend like it never happened.

No one's gonna know this is you.

This isn't the a lead based on the interrogation.

Even if you don't have that actual mood, you have to sorta put it on.

You have to accept that he's president.

He could come up here at any time.

I can't let you risk that heartache.

You get all four wheels spinning in the same direction.

Well, you know what? Why would I want to?

Then we don't have anything to worry about, do we?

I get that, but you sure it's not something more?

It's a recording of conversation that took place four months ago.

My assistant will answer any questions you might have.

Why such a controversial subject in the short form?

So we went to Anna's house, and that's when I saw you two through the window.

Their phrase for advertising to women was always a bridesmaid, never a bride.

I'm the guy who called a little while ago.

Look, the police are going to be coming back.

No no no. Can't leave him like that.

Careful, that stuff is slippery.

I don't remember going deep in that story like I did.

What a perfect guest tonight for the Christmas season!

Usually it's New York, but this time he said Miami.

Did you really mean it when you said that you thought I looked beautiful? Because I can't talk if I'm unconscious.

What's wrong? I already computed this inverse, and now it's not working?

I hope this is not the last time I speak to you.

Well, is it just us? I mean, is it only happening to us?

She's giving up something she worked really hard on, so you will understand that you're her first priority.

He has to go back to the future because he's in the past.

Because at night time there's not many people on there.

You know what? Just give me the dinosaur one.

We will take some back roads to see some interesting stuff out here.

It took me half a day to work that out mathematically.

I'm going to need to let it dry.

Well, today, for the first time, I'm gonna say no.

And its goal very possibly being a very important landmark on or near this coast.

She grew up here in New York City.

Because we're going to learn to be happy because we deserve to be.

He sees himself primarily as a musician.

Without surgery, condition could be fatal?

He is not supposed to go on those anymore.

They're basically the message they're sending us, "Just do something".

But if I want to succeed, I have to keep trying new things, finding new talent.

Really? Well, I'm not sure they were the pilots.

I'm human, you know, like everyone else.

The day that you graduated from the police academy, honey, that was the proudest day of my life.

But if not, it's not healthy for the children.

Now look at this, only one minute farther down the road.

They need to become somewhat self-sufficient.

What really matters? What's good for me? What lasts a lifetime?

I knew that you would come to your senses.

Hey look, you know what? Victor has a huge house.

Say we're pleased to see them and thank you for coming.

The time frame is absolutely perfect because the sublet's up on February 1st.

His BP and heart rates are up a little, nothing too worrisome.

That having been said, there are many reason why this happened, as I understand it from several sources.

I'm going to blow up the face of everybody in that crowd, and I'm going to find him.

She told me she was going off this year.

I can only promise you, when I become rich in America.

Well, we're just very lucky to be in the presence of who no longer takes, but gives.

I don't think there's anything further we have to say to each other.

Everyone has their own style and their own comfort level.

Ok, but I'm going to pay you back, though, ok?

I think if you just gave him a chance you would find him an outstanding officer.

I'm sure those guys get a lot of attention from women. They travel all over the world.

If you're gonna trust me, you're gonna come to it in your own time.

I made an undetermined amount of money that was not their money.

I'm tired of being reduced to a faceless resume on some editor's desk.

I wanted to, you know, make the commitment and let my son know that I'm serious.

We decided to try to be on time for a change.

You better scram before the boss breaks your legs.

I didn't read it that way at all.

And your husband won't like your construction crew.

I've got very big news that concerns you.

I want to reclaim what's mine, leave no doubt in anyone's mind that the court is making the right decision.

I made bean salad for us for lunch.

I again apologize for my misbehaviour.

We would have to tell Taylor what test we were running and why.

Thank you for all the hugs you gave me.

I had to deal with that pain, pushing it away and pretending it didn't matter.

Just the idea of food prison bothers me.

Most women can go a lifetime without a moment like that.

Something other than what I'm supposed to do.

Thank you for being supportive, for being so strong.

Rejected? Now you're going to see what it's like.

Yesterday, she said the piece was in the safe.

I learned that the hard way.

I don't think " convenient " is the right word here .

It wouldn't surprise me if she skipped down.

Although I conceded the election, I am still the district attorney.

Many white-collar employees will be offered buyouts.

You can always do the fun stuff.

I give the order, you hit and don't miss.

You know Teresa is a very sensitive person.

I would like to have it on a business basis with somebody.

It was one of the more memorable shows for the three of us.

There were these blank spaces and I was hoping that you could help me fill in the gaps.

I mean, I don't buy her ponies, but I'm still her favorite uncle.

A few years back I had a little cash flow problem and was forced to sell Channel.

They're redoing the pharmacy there at the Walmart.

My grandmother told me you were taking care of Lulu.

And besides, I need to be in a city where I can get a good Texas meal.

The study says that chronic inflammation can negatively affect brain chemistry.

My dad won't admit that booze was a factor, that he didn't have any business getting behind the wheel.

So what? All you had to do was say something and we would have let you go.

All right then, now it's time for you to stick to your side of the bargain.

That's what everybody perceived it.

That will guarantee me a seat on the board of directors.

We can probably catch a movie or go to that art gallery or something.

What they don't know is how amazingly short I can make a toast.

The most exciting plans for this summer to bring french canada and english canada to-

gether.

Why don't you just stay here with your family, and I will call you tomorrow?

So he got another radio this summer, but of course that got ripped off.

The american dental association convention is this week, But as soon as I get back, I will look into it.

It's not the size of the building or the size of the ad, it's the determination and the passion put into the fight.

I didn't write about it until I had someone on the record talking about it.

Not on the line? Makes no difference.

So for little Ben and his older sister, I made one just for you.

I know that you're also writing a more light-hearted approach to aging that's a little less intense.

Talk just like we're old buddies.

Bart station around ten o'clock?

Pat and I will proliferate in a moment.

Perhaps you would consider taking rooms here at the castle.

The last thing I remember was leaving your mom's house and going to look for her.

Usually she doesn't stay up too late.

You walk in every kitchen of every hotel in America.

You're doing a great job ruining everything all on your own.

Mother is taking a break before their children came home from school.

I thought it would take weeks and weeks of battles and twists and turns.

I'm certainly not going to mention it if he doesn't mention it.

So winning today's challenge and its prize of immunity means more than ever.

This one day, just before Christmas, one of the doctors comes in, looking all serious.

One of the reasons Chloe went for me in the first place was that I knew how to have fun.

She can sing at the wedding, but the honeymoon is off.

You think this is the best way to get a job here?

I'm going to take down Gary and Jane.

That's hilarious.

In order to stand on your feet, you have to do things which you feel today are not so pleasant.

I come up with my own ideas about that stuff.

But, anyway, I'm using up all the tape here.

And thank you for sticking by me and helping me get through this.

Do it? Do it? Teach them properly, Major.

I think that we have gotten through that at this point.

As I tried to explain before, you can not get honey from a hornet's nest.

We are planning this wonderful trip to Hawaii, we're going to spend some quality family

time together.

Can you believe it? I think I've reread the same page for the last ten minutes.

Right off the bat, they don't like guys like us.

Who are we? Where have you been? This requires everyone.

Go talk to him before somebody else dates him.

Anyway, I'm going to wait here until he is stabilized, and then I am leaving.

I bit my tongue the other day.

Ain't that right, Joe? Saw her head into Doctor's office.

There could have been some clean-up crews around.

These rings are a symbol of all of love's unbroken circle.

We were in your apartment together.

Like when you used to play the board games with the kids in front of the fire.

I thought we had an agreement, that we were gonna give that money to where.

What kind of ice cream was that?

I just can't believe that he doesn't know you well enough to know that you would fight back.

But we kind of give people permission to challenge conventional wisdom sometimes.

Drug abuse is now the number one cause of accidental death, according to the federal government.

No, actually, it can wait till later.

Your grandpa knows that we are up here in the attic, but no one else can know we're hiding here.

You know that? You're the one who should be getting angry, and you're not.

Can't you reschedule or something? Sorry, man.

It's like a game of chess, where he knows what moves you're going to make before you even make them.

They're not even going to notice I'm gone with everything that's been going on.

It was funny, it was entertaining, and it ripped my heart out.

Surely they must know each other.

She said that they will do anything to keep me from getting into that office.

If anything, it's the other way around.

It's different things in different places.

They were held up, so they're coming in the morning to do all the renovations in here.

I said if you invite them up there for a party, they're gonna assume that they are staying with you, right?

Do you want to play the hero? It's time to live up to your word.

I don't think that I've seen as many of you in one place since they announced the results of my first bar exam.

I'm not done yet, but I don't know where they are.

We got to get it done right.

But we're going to look at it very carefully when the results are all in.

I've got stuff growing out of my locker, too, but I think it's mold from an avocado sandwich or something.

I mean, odds are, you're never going to find out who your biological parents are now, anyway.

Why did you let me wear these shoes this morning?

Go with my uncle's plan B for the money.

Can I grow some basil?

But the fact of the matter is, this is the setting for the greatest story ever told.

If they need stitches, they can go to you.

No particular emotional investment in this unfortunate transaction.

Why are you wearing your mask?

I said, you have to do me a favor, we don't want these big deficits.

So we are going to stay right here, and figure out a way to make the money come to us.

It's not like we can squeeze through the slot drawer.

You notice anything different about Carrie, boss?

We're going to leave all this and we're going to buy the best red cupcakes we can find.

Is this your first appointment?

I still remember when you were afraid that corner store guy didn't like you.

You do not have to have every single piece of the puzzle to determine the picture.

Can you tell me where you were that night?

What picture? I don't know what you're talking about.

You might be the prettiest one yet.

Except that he thinks he saw it for real.

And finally he got to the other side, and the boy told him who he was.

Look, I'm only going to be a few minutes, so park it up front for me, would you, please?

What else you suppose to do? Surprise you took it so long.

No, you just need to get a good night's rest.

Will you pass me some of that tea please?

I felt like we had a lot in common, like I could hang out with him.

All of the week's politics are on our roundtable.

I just wanted to stop by and drop off this letter.

That's all part of the game, boy.

Do you all wanna see a little bit of this? Okay.

And certainly, it is an issue that has to be addressed.

That is what's going to stop her.

We will be scared and we will be freaked out, but at least we will be in this together.

Is there anything else I could take on in addition to my regular duties, sir?

Those are the stem cells we have those organs and those tissues regenerated when things go wrong.

We get up in the morning, sometimes we feel good, sometimes bad, but we go through it with feeling.

Do you want it? You should never say no to a gift.

I'm not letting Sheridan stay the night with Antonio.

She's been like this for as long as I can remember.

I don't think that this set Donald Trump off his game or anything like that.

It's not exactly Disney World, is it? And the storm's coming.

There must have been something of a rivalry.

That's him in disguise, you know.

It makes absolutely no sense.

And brush your teeth too? He'll be here any minute.

I'm gonna have a couple of drinks.