Textflow: Towards Supporting Screen-free Manipulation of Situation-Relevant Smart Messages

PEGAH KARIMI, EMANUELE PLEBANI, AQUEASHA MARTIN-HAMMOND, and DAVIDE BOLCHINI, Indiana University-Purdue University Indianapolis, USA

Texting relies on screen-centric prompts designed for sighted users, still posing significant barriers to people who are blind and visually impaired (BVI). Can we re-imagine texting untethered from a visual display? In an interview study, 20 BVI adults shared situations surrounding their texting practices, recurrent topics of conversations, and challenges. Informed by these insights, we introduce *TextFlow*: a mixed-initiative context-aware system that generates entirely auditory message options relevant to the users' location, activity, and time of the day. Users can browse and select suggested aural messages using finger-taps supported by an off-the-shelf finger-worn device, without having to hold or attend to a mobile screen. In an evaluative study, 10 BVI participants successfully interacted with *TextFlow* to browse and send messages in screen-free mode. The experiential response of the users shed light on the importance of bypassing the phone and accessing rapidly controllable messages at their fingertips while preserving privacy and accuracy with respect to speech or screen-based input. We discuss how non-visual access to proactive, contextual messaging can support the blind in a variety of daily scenarios.

CCS Concepts: • **Information Interfaces and Presentation → Assistive technologies**.

Additional Key Words and Phrases: Text entry, Assistive technologies; Intelligent wearable and mobile interfaces; Aural navigation, and Ubiquitous smart environments

1 INTRODUCTION

Mobile texting mainly relies on on-screen keyboards that display characters visually, posing barriers to people who are blind and visually impaired (BVI) [2, 52]. Even when keyboards have accessible overlays (such as VoiceOver), they remain screen-centric: manipulating text is generally bound to holding a device out at all times and interacting with visual keypads primarily designed for sighted users, or with accessible keyboards that read aloud keys upon touching [16, 28, 32, 55, 57]. Screen-bound interaction is especially problematic in nomadic contexts, when blind users are on-the-go, have to keep a cane in one hand and the other hand free to touch objects nearby. Users can resort to voice input [6] but are still often required to be bound to a mobile device for dictation and go through several repetitions to overcome recognition problems and ambient noise [5]. Prior studies suggest users who are blind tend to prefer not to use speech-based interfaces in public places, due to concerns for privacy and social conspicuousness [1]. Another major issue of current mobile messaging is that, whereas daily routines and recurrent situations often lead users to periodically re-send similar messages, current approaches always require users to initiate an open-ended text composition. As a consequence, there is an undue burden to blind users, especially in light of the severe mechanical and interaction constraints they have to overcome to complete a text. Can we re-imagine entirely non-visual texting in a way that is quiet, proactive, and untethered from a screen? To explore this question, the paper makes three main contributions:

Authors' address: Pegah Karimi, pekarimi@iu.edu; Emanuele Plebani, eplebani@iu.edu; Aqueasha Martin-Hammond, aqumartin@iu.edu; Davide Bolchini, dbolchin@iu.edu, Indiana University-Purdue University Indianapolis, Indianapolis, Indiana, USA.

2160-6455/2022/1-ART1 $15.00
https://doi.org/10.1145/3519263

- A dataset of more than 850 parametric messages (17 topics as labels and >250k instances of texts) derived from an existing public dataset called Weibo [25] that contains short text conversations. Topics are informed by the results of a study conducted with 20 BVI users that determined the type of text messages BVIs send while on the go or in public places. Furthermore, we associate the texts contained in the dataset to recurring situations and daily activities in which BVIs send text messages while outside the home, providing insights into specific topics most relevant in each context of use.
- TextFlow: a mixed-initiative context-aware system that generates entirely auditory message options potentially relevant to the user situation. TextFlow enables users to listen to, browse and send suggested aural messages without holding or attending to a mobile screen, using nimble finger-taps supported by an off-the-shelf finger-worn device. The system operates as a sequential auditory stream of fast-spoken topics and messages that can be browsed and selected for composing a text.
- An evaluative study with 10 blind participants who interacted with TextFlow through 15 tasks focused on daily text messaging. With an average task success rate of 88.6%, all participants were able to interact with TextFlow and send messages in entirely screenless mode. An in-depth analysis of the user performance shows that the error rate is lower when participants scan the system's suggestions to make a selection as compared to when they listen to a self-disclosing auditory stream of items. In addition, as participants interact more with the system, its learnability quickly improves. Participants consistently appreciated the positive experience of bypassing screen-centric methods for typing or voice input, especially when they are mobile and need to send short notifications without taking the phone out of their pocket.

2 RELATED WORK

2.1 Eyes-Free Mobile Text Input

Navigating mobile keyboards using touchscreen devices has been widely adopted by blind users. Technologies such as Apple's VoiceOver [37] allow users to touch the screen and get feedback via voice. While voiceover interaction makes mobile keyboards accessible, it requires users to continuously move the finger on the screen to find and select the intended letter. Researchers introduced better typing methods by rearranging the alphabets of mobile keypads. BrailleTap [27] and NavTap [28], for example, eliminate the need to memorize the letter position by grouping 3 or 4 letters as keys. Other technologies combine multi-touch input with audio feedback to enable fast typing. No Look Notes [11] is a multi-touch text entry that arranges letters in pie menus. Users can tap their fingers on the screen to switch to the desired pie menu. Additionally, Escape Keyboards [8] allows users to keep the phone in one hand and use their thumb to press letters on the screen by performing a flick gesture. Multimodal Text Input Touchscreen Keypad (MTITK) [16] is another example of multimodal text interaction that divides the screen into a 4x3 grid and groups letters into five keys with distinct audio-tactile feedback. Users can touch the screen to recognize the intended letter, tap to make a selection, and use a gesture to edit the text. These approaches enhance the accessibility of typing, but are still screen-centric, thus requiring the blind to hold out a phone at all times to interact with a visual display.

Speech input is now broadly available and enables voice dictation for texting [6]. When used outside, however, besides causing issues when background noise is present [5], speech input also introduces barriers to effective use. For example, it is challenging for the blind to review and edit the dictated text on the screen [6]. In addition, recognition accuracy is often reduced when background noise is present [5]; thus, requiring the user to repeat the text multiple times. Social and privacy boundaries can also be easily broken by voice input, and this is problematic especially when the blind prefers not to draw unwanted attention in public spaces [17, 56].

Wearable devices including armbands, hand- and finger-worn devices open additional input channels that can potentially bypass the constant reliance on a reference screen. NailO [34], for example, is a nail-mounted wearable surface that detects various swipe gestures to input emoticon or punctuation without the need to use the screen. However, users still have to rely on the mobile screen due to the difficulty of typing using only the nail surface. Recent work has shown the use of the Myo [49], a hand-worn band that recognizes and uses as input the arm's muscle movement, to control entirely auditory keyboards for the blind without the need for a reference screen [46]. The approach still requires performing quite ample hand gestures that are not as discreet as users would like, and may unnecessarily fatigue the arm.

2.2 Facilitate Texting using Machine Learning Models

Existing intelligent interactive systems leverage advances in natural language processing (NLP) and artificial intelligence (AI) to facilitate texting by predicting the next phrase while the user is typing [24, 38]. These approaches have resulted in features, such as message autocompletion or autocorrection, which are broadly available on mobile platforms. However, most of these algorithms have two main limitations: first, they do not link phrases to users' context, such as time and location. Therefore, they do not take into account users' intent in a given situation. Second, they fail to make accurate and relevant predictions without the user entering most of the words of a sentence [33]. This is because machine learning models integrated into these systems [18, 38, 54] use examples of texts from popular media such as Twitter [44] or Reddit [58] that contain one-to-many communications instead of interpersonal text conversations. Interpersonal text communication tools, such as iMessage or WhatsApp, do not have publicly available data due to users' consent and privacy. Question-answering datasets also include samples of short sentences. An example is the SQuAD dataset [30] that contains around 100,000 question-answer pairs on more than 500 articles. Another example is Natural Questions [36], a collection of Google queries with annotated short and long answers. However, these datasets focus on retrieving knowledge from external sources (e.g., Wikipedia) and thus fail to incorporate interpersonal text conversations. The closest dataset we could find is Weibo [59], which contains short text conversations but does not provide annotations about users' intent or plans. In addition, most samples from Weibo cover a wider range of communication (e.g., advertisement). As a result, most of the existing public datasets are not specific enough for one-to-one text communication and do not provide annotations for user intent and context. To address the shortcoming of existing public benchmarks, this paper introduces a large-scale dataset of short text conversations based on a study conducted with 20 BVIs about their needs and practices while on the go.

2.3 Context-Aware Mobile Communication

Research in ubicomp has studied the potential of leveraging contextual dimensions, such as location and time, to facilitate human communication through text messages. *Location* is a common type of contextual cue deployed in ubicomp applications [3]. Perry and Shangar [53] address the importance of the physical location in space in human communication and argue that it is common for people to stick notes on their desks as a reminder or put relevant notes in social areas. LAMMS [10], for example, demonstrates the notion of location-based text messaging, enabling people to communicate locally with each other and receive messages related to that area.

Time is another dimension often used when sending text messages. LATTE [51], for example, is an email-based system that incorporates both time and location to dynamically expand emails with the corresponding temporal and spatial information. Other forms of context-based mobile messaging systems allow users to define the context while sending a text. Jones and Neil [31] introduced a contextual constraints model, in which the user can place constraints on whom to send the message to (i.e. personal dimension), where the message has to be sent (i.e. spatial dimension), and when the message has to be sent (i.e. temporal dimension).

Contextual information has also been used to provide feedback about the environment while an individual is on the go and holding the phone for typing [48]. Many people experience walking and typing at the same time they often fail to perform both tasks perfectly [47]. One method for providing feedback about the user's ambient environment is through audio [15, 42]. These systems provide real-time feedback to users allowing them to focus on typing while reducing errors by increasing awareness about the environment. However, the method can create difficulties for BVI users due to receiving auditory information from both typing and their ambient environment.

2.4 Use of Context in Recommender Systems

Context-aware recommender systems use the concept of ‘context’ to recommend items that are both relevant to the user’s preferences and their specific context such as time and location [4, 7]. Magitti [9], for example, is a context-aware mobile recommender system that automatically detects a user’s activity based on the user’s context and behavior to recommend personal and timely leisure activities associated with a local environment. SoCo [40] is a social network-aided recommender system that combines location and time with social information learned from friends with similar tastes to provide highly accurate recommendations. Bouidghaghen et al. [12] introduce a situation-aware recommender system to alleviate the problem of information overload. The system retrieves contextual information related to the user’s location and time and provides a mobile user with personalized search results. However, context-aware recommender systems that rely on past experiences (i.e. exploitation) cannot model user’s interest evolution because the learned rules will only reflect past user behavior; for this reason, a fraction of recommendations are selected at random or using a heuristic to obtain information about the user and discover better recommendations (i.e. exploration). In [39] authors use a bandit algorithm to exploit well-established advertisements for short-term capitalization as well as exploring less-known cases to find out potential advertisements that can be recommended in the future. Bouneffouf et al. [14] introduce a hybrid-ϵ-greedy algorithm that takes into account the context to deliver recommendations that are bound to the user’s current situation and interests. The algorithm combines the user’s situation based on time, location, and social ontologies with an ϵ-greedy algorithm and content-based filtering techniques.

3 ELICITING MOBILE TEXTING NEEDS AND PRACTICES

We conducted a formative, interview-based (IRB-approved) study to better understand the practice of text messaging adopted by blind and visually impaired individuals. The goal of our user study is twofold: 1) elicit recurrent messaging topics associated with BVI specific needs for mobile text communication; 2) explore the types of daily situations where BVI engage in text-based messaging.

3.1 Participants and Procedure

The study engaged 20 BVI participants (12 females and 8 males). Eight participants identified as totally blind, four legally blind, five had minimal light perception, and three had low vision. Participants’ ages ranged from 35-79 (μ=51.7, σ=11.2) years old. 19 participants used iOS devices and one used Android. The average length of mobile usage was 11 years. We conducted the interviews through phone calls or Zoom meetings, depending on the participant’s preference. Each interview lasted approximately 1 hour. Interview questions covered three foci: (1) places participants typically visit during daily routines and modes of transportation; (2) situations and circumstances in which they might send a text; (3) recurrent topics of their texts and most recent text messages they sent to others. Interviews were audio-recorded for transcription and analysis. Upon completing the study, each participant received a $30 Amazon gift card via email for their participation.

3.2 Analysis and Results

We performed a thematic analysis of the participants’ responses. Overall, six high-level themes emerged that represent different types of text messages participants send on a regular basis: *notifying someone*; *offering assistance*; *scheduling/rescheduling plans*; *coordinating with someone*; *reminding someone*; *requesting assistance.*

3.2.1 ***Notifying someone***. Participants send different types of notifications according to their immediate situation. One such notification is to let others know that they might *arrive late* or specify their *arrival time*. For instance, P1 commented: "*If I'm running late to work, I may send a message to my boss saying, hey, I'm running late, I'll be there in 20 minutes.*" In other instances, participants send a text to let others know they are *on their way*. For example, P7 would say "*The bus is ready to come, I am on my way.*" when going to an appointment to meet a friend. Participants mentioned that when they are on their way to a specific location, it is common to inform their close ones that they arrived safely or to *signal their location*. For example, P3 described the type of message they would send when they are outside: "*I send a text to my husband and say I got here okay or I'll tell him that I got to the bus stop safely.*" In some cases, users might *not be able to talk* due to privacy concerns and therefore might send a text stating that they will call back later. P2 gave an example of when they are waiting at the bus stop: "*Let me get back with you later, I am in the middle of commute.*" Another typical notification that participants mentioned was to let others know that their initial *plan has changed*. For instance, P6 described their experience of sending a text while they were on their way to work, stating: "*while waiting at the bus stop, I was notified that our school trip that was supposed to go to Chicago got canceled, and so I had to text to colleagues that the plan for that day has changed.*" On some occasions, participants specified that they had to send a text to emphasize that they are *waiting*. P11 gave an example of a text message they would send when they are waiting at the bus stop: "*are you almost here? I am waiting for a long time.*" Other, less frequent topics included letting others know they are in a *meeting* and on their way to go *shopping*.

3.2.2 ***Requesting assistance.*** Participants may send a text asking for help due to access barriers, in particular when they need to *get to a location*. Most participants mentioned that when they reach a destination or they are inside a building, it is often difficult for them to find the entrance or exit. In these situations, they send a text to the person they planned to meet to look for them and guide them to the place. For instance, P2 explained an experience, stating: "*One time I was in a building, and I couldn't find my way. I was on my way, but I could not find my way back because I did not have my GPS on. So, I sent a text, and I said I got lost, and then I asked for help.*" P8 described the type of text message they would send if they could not find the entrance: "*Come outside and get me.*" Another type of request that participants mentioned is to ask someone to be *picked up*. P15 gave an example of the type of text message they typically send while walking to reach the bus stop: "*pick me up in five minutes.*" In a few cases, participants mentioned they might ask someone to *bring food* when they are at work.

3.2.3 ***Scheduling/rescheduling plans.*** When on the go or at work, it is common among BVI users to send texts to *make or change plans* for both formal and informal commitments. For instance, P5 provided an example of a text they would send to a friend when they are at work:"*where would you like to go for dinner?*" The same participant gave an example of a text sent on their way to meet a friend: "*I'll see you at 2 for our appointment.*" Users may also need to reschedule existing plans. P10 gave an example of a message they would send when at work: "*9'o clock meeting is canceled; meeting is at 10.*" Participants also described sending a text to friends or family members about a change of plans. P16 explained a text sent to a friend when they realized they couldn't meet as planned: "*I told him I had the flu and I wouldn't be able to make it. And then I said could we reschedule?*" On other occasions, users might send a text to schedule a *phone call*. P3 gave an example of a text they would send on their way to work: "*when can we set up a call?*" Thus, BVIs need to be able to communicate their plans quickly while on the go or at work.

3.2.4 ***Coordinating with someone.*** When on the go or in public places, users often need to text others to coordinate about location, time, and the number of people they expect for a gathering. For instance, P1 gave an example of a text message they sent when meeting a friend at a restaurant: "*are you here at the restaurant yet?*" The participant is trying to *check-in location* to find out whether their friend has arrived at the restaurant or not. P9 gave an example of coordinating a group meeting: "*if we're supposed to be like a group of people, I can't remember everybody who's supposed to be there. So, I will send a text and say: who's going to be for dinner?*" The participant is *checking-in guests*, as they may not have instant access to a phone to verify the number of friends who confirmed coming for dinner. When asked about the context of the messages sent while returning from work, P15 stated: "*Since everybody is on a different schedule, we mostly text and say what time you're going to be home for dinner.*"

3.2.5 ***Reminding someone.*** Unlike sighted users, BVIs experience more difficulties accessing their calendars to check their appointments. In these situations, they often send each other reminders to confirm the upcoming *appointment* or *meeting*. P17 provided an example of a message sent to a friend while walking to the bus stop: "*See you tonight at dinner.*" When at work, participants might send reminders related to meetings or ask for confirmation. P10 provided an example of a text message they might send to a colleague: "*is the meeting today at 9:30?*" On other occasions, users might send a reminder related to an upcoming meeting to verify if others will be there. P13 commented: "*our meeting is today at 11 o'clock, please let me know if you can attend.*" Therefore, users send reminders to confirm meetings and/or appointments both for themselves and others.

3.2.6 ***Offering assistance.*** When on the way, users offer assistance to others regarding daily activities. For instance, P5 stated: "*If I'm in a vehicle, I have [to] kind of hover a lot to send a text to my son and say, do you want something specific from [the] drugstore?*" The participant emphasizes the amount of time it can take to send a text to someone and ask a question related to *shopping*. Users might also be in a specific store and send a text to ask if the recipient needs anything from that store. P15 gave an example of this type of text message: "*I'm here at the dollar store. Do you need anything?*" Both participants in these examples are offering help to avoid an extra trip to go to a store. Other types of offering assistance include *picking up* someone on the way or helping to *get to a location*. P15 explained their experience when they have an appointment with a friend, stating: "*I might text my friend and say, do I need to meet you at the bus stop to help guide you in?*" Therefore, BVIs frequently offer assistance to each other and members of their household while outside the home.

In total, 17 specific topics emerged based on our study: *arrive late*, *arrival time*, *on my way*, *get to a location*, *check-in location*, *check-in guest*, *signal my location*, *appointment*, *meeting*, *shopping*, *making plan*, *changing plan*, *pick up*, *waiting*, *phone call*, *not able to talk*, and *bring food*. In addition, we identified specific situations in which some of these topics emerged frequently. Figure 1 shows the occurrence of topics that appeared three or more times for four different situations.

3.3 Generating Messages from Elicited Topics

For each topic, we generated a list of short candidate messages. For instance, if the topic is 'arrive late', a potential message would be 'I will arrive a few minutes late.' To generate messages, we followed three main steps: 1) defining initial candidate message types that are derived from the study with 20 BVIs; 2) augmenting message samples using an AI model that retrieves similar samples from an existing public benchmark, and 3) determining parametric message instances based on factors such as time and subject of the conversation, and then using a word prediction algorithm to generate different values for a given parametric message instance. Figure 2 shows the three steps and message constructs defined for the message generation. The example underneath each stage refers to the topic 'arrive late'.

3.3.1 ***Defining candidate message types.*** We intersected topics and messaging actions that led to one or more candidate message types, allowing us to generate message samples (step 1, Figure 2). For instance, the intersection of *notifying* and *arrive late* leads to a potential candidate message of 'notifying others that you/someone will be late'. We defined candidate message types based on text samples provided by participants (a total of 117 samples) as well as manually intersected topics and messaging actions to define potential candidate message types. The result of the intersection led to 93 candidate messages and 180 message samples. Table 1 shows the intersection of topics and messaging actions, in which the intersection that is based on the user study is shown in bold and the manual intersection is left in regular text.

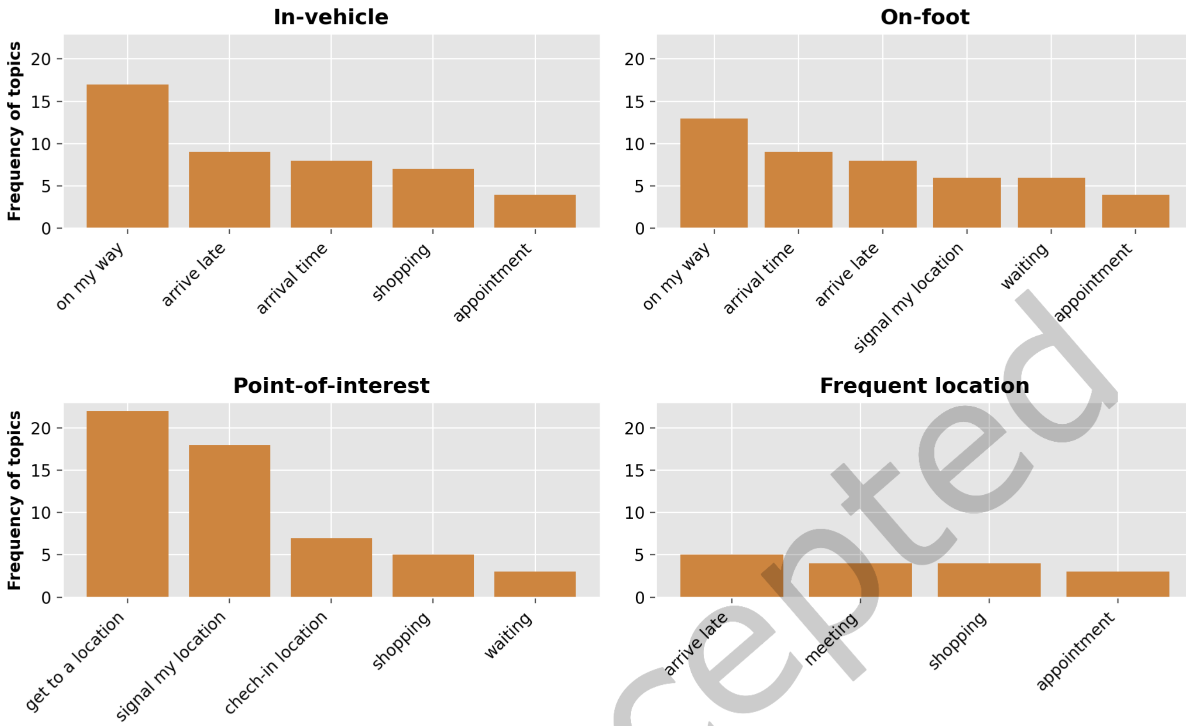


Fig. 1. The frequency of topics identified by participants as relevant to four different situations: in-vehicle, traveling on-foot, nearby or inside points of interest, and when going to a frequently visited location (e.g. workplace).

3.3.2 ***Augmenting message samples.*** To augment the message samples associated with each candidate message type, we used a public benchmark called Weibo [59] that contains interpersonal short text conversations from a Chinese microblog service. Weibo has more than 4.4 million samples, divided into posts and comments associated with each post. We selected this publicly available dataset because it contains a very large number of instances of the specific textual genre we are seeking to model, rather than broadcast, one-to-many communications typical of social media platforms. Samples are translated from Chinese to English using a tool called Youdao [60]. We performed data cleaning by removing redundant words such as 'come on' or 'oh' and discarding generic topics such as politics or advertisement. Given a topic, we used a pre-trained language model called RoBERTa [41] to retrieve related sentences from the dataset (step 2, Figure 2). The RoBERTa model has 24 transformer layers with a feature-vector size of 1024 and 16 attention heads. The model has been trained on the Multi-Genre Natural Language Inference (MultiNLI) corpus [50] with pairs of sentences annotated as entailed, contradictory or neutral. For each topic, a reference text sample was selected from the user study. If the topic belongs to more than one category of text message, a sample was selected from each category. For example, if the topic 'appointment' belongs to both reminding someone and coordinating with someone, we selected a reference sample from both categories. We compared the reference sample with all the sentences in the dataset using pairwise sentence classification. We then retrieved the sentences classified as "entailment", i.e., the meaning of the query sentence is implied by the retrieved sentences, and "contradiction", i.e., the retrieved sentences imply the negation of the query sentence. The contradiction class was used only when appropriate to the message sample. For instance, for

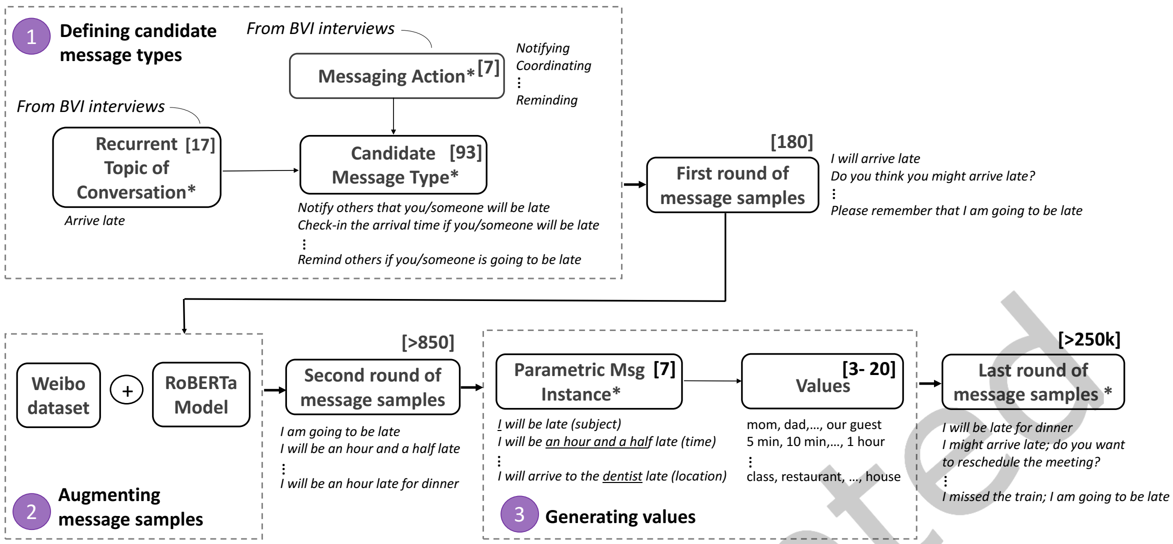


Fig. 2. The conceptual model used to generate message samples for BVIs while texting on the go. The number in parenthesis indicates the number of instances of each construct for each data generation stage.

the topic 'arrive late' and message sample of 'I will arrive late' we did not use the contradiction class because the retrieved samples, such as 'I will arrive earlier' or 'I will be early for dinner' would change the purpose of the topic. The model returns a probability for each class as a function of the sentence pair:

$$(p_c, p_n, p_e) = f(s_1, s_2)$$

where s_1 and s_2 are the two compared sentences and p_c,p_n,p_e are the probability of contradiction, neutral and entailment respectively. We rank the results by p_e and p_c, i.e. the entailment and contradiction confidence scores. For entailment, we consider sentences with p_e higher than 0.2 (20%), whereas for contradiction we consider only sentences with p_c greater than 0.1 (10%). This choice is motivated by the fact that the entailment class provides more diverse and related sentences for the given query, whereas the retrieved sentences for the contradiction class are limited to a few samples. Low-scoring sentences, i.e. below 20% for entailment and 10% for contradiction, are not considered as they are often unrelated to the query.

Starting from the first retrieved recommendation, we selected those that included the topic and a single instance of contextual information based on factors such as time, location, audience (e.g., with someone), purpose (e.g., for dinner), or method (e.g., with Uber). For instance, the sentence "I've been waiting for a long time" consists of our topic (waiting) and information that concerns time. Recommendations that were not relevant to our topic or that had the same contextual information as the selected sentences are skipped. The third column in Table 2 shows examples of the selected recommendations for the topic 'waiting'. Selected recommendations were then refined to represent a short message (fourth column of Table 2). The steps for refining sentences are the following: 1) we added or modified the subject to be in the first person; 2) recommendations that included more than one sentence were refined by removing sentences that did not contain the keyword related to the topic; 3) selected sentences that referred to specific locations or times were replaced by either the user's current situation or a plan elicited from the calendar. After augmenting message samples, we were able to obtain 856 text messages.

Table 1. Intersecting topics with messaging actions to generate candidate message types. The "Bold" means the intersection comes from the user study with 20 BVIs, whereas the regular text is based on manual intersection performed by researchers.

	Messaging Actions						
	Notifying	*Requesting help*	*Offering help*	*Coordinating*	*Scheduling*	*Rescheduling*	*Reminding*
Recurrent Topics	**Arrive late**	Arrive late	Arrive late	Arrive late	Arrive late	Arrive late	Arrive late
	Arrival time	Arrival time	Arrival time	**Arrival time**	Arrival time	Arrival time	Arrival time
	Waiting	Waiting	Waiting	Waiting	Waiting	Waiting	Waiting
	Shopping	Shopping	**Shopping**	Shopping	Shopping	Shopping	Shopping
	Appointment	Appointment	Appointment	Appointment	**Appointment**	**Appointment**	**Appointment**
	Meeting	Meeting	Meeting	Meeting	**Meeting**	**Meeting**	**Meeting**
	Pick up	**Pick up**	**Pick up**	Pick up	Pick up	Pick up	Pick up
	On my way	On my way	On my way	On my way	On my way	On my way	
	Signal my location	Signal my location	Signal my location	Signal my location	Signal my location	Signal my location	
	Bring food	**Bring food**		Bring food	Bring food	Bring food	Bring food
	Not able to talk			Not able to talk	Not able to talk	Not able to talk	Not able to talk
	Phone call			Phone call	**Phone call**	Phone call	Phone call
	Get to a location	**Get to a location**	**Get to a location**	Get to a location	Get to a location		
		Check-in location	Check-in location	**Check-in location**			
				Check-in guest			Check-in guest
	Making plans				**Making plans**		
	Changing plans					**Changing plans**	

3.3.3 ***Generating values with message samples.*** Our last step to further augment the dataset is determining parametric message instances for potential message samples. Parametric message instances are based on seven factors: time, mealtime, meal, forms of transportation, social events, location, and subject. These factors correspond to different values within a sentence (step 3, Figure2). For instance, within a message sample of 'I will be late for dinner', 'I' and 'dinner' are parametric message instances and can be replaced by other values that represent subject and mealtime. To determine alternative values, we used the same AI algorithm (RoBERTa) in order to predict word(s) within a context for different message instances. The model ranked the predicted values based on their probability score, allowing us to consider only the top 30 predictions. We performed a manual inspection for each prediction and selected values that are suitable within a message sample. For example, if the message sample is "I will wait at the gate" with the word "gate" as the parametric message instance, the AI model predicts the following words to replace "gate": door, airport, end, bridge, alter, bottom, border, etc. In this case, we will discard values such as 'end', 'alter', and 'bottom'. When the factor is time, we put a five-minute increment to represent a regular period. We kept the message samples provided by participants (117 samples) the same and applied word prediction for candidate message types and those retrieved from the Weibo dataset. The number of values for the selected sentences falls between 3 to 20. Table 3 shows examples of parametric message instances (shown with an underline) for the topic 'arrive late' and 'bring food' along with the predicted values after manual inspection.

Table 2. Example of messages for the topic 'waiting' for entailment and contradiction classes.

Reference Sample	Class Type	Selected Recommendations	Messages
I am still **waiting**	Entailment	I am waiting for you to come	I am waiting for you to come
		I've been waiting for a long time! where have you been?	I've been waiting for a long time!
		I am waiting at the dentist	I am waiting at *
		I am still waiting for someone	I am still waiting for someone
		Wait a few minutes!	Wait a few minutes
		Waiting outside the maternity ward	I am waiting outside the *
		Waiting for dinner!	I am waiting for **
		Still waiting for my ride	I am still waiting for my ride
		Wait a few seconds	Wait a few seconds
	Contradiction	There is no need to wait for this photo	There is no need to wait for me
		I am not waiting anymore	I am not waiting anymore

* current location based on the map. ** breakfast, lunch, or dinner based on the time.

4 AUGMENTING MESSAGE SAMPLES WITH SITUATIONS

Texting is a situated activity. People text in various contexts when they are on the go or in public places. Based on the result of the initial study, we characterize four different situations in which users send a text:

(1) **In-vehicle:** This situation describes context in which a user is in a car, bus or transportation vehicle.
(2) **On-foot:** The on-foot situation refers to when BVIs are walking or waiting while en route to bus/train stations. This situation is broken into walking to stations and waiting at stations.
(3) **Point-of-interest:** Point-of-interest (POI) refers to noted places that include public and private places (e.g., schools, restaurants). The situation is divided into two categories of indoor points-of-interest and outdoor points-of-interest in close proximity.
(4) **Frequently visited location:** Frequently visited location (FVL) defines places that BVIs go on a daily routine, such as a workplace. This situation is broken into indoor frequently visited locations and outdoor frequently visited locations in close proximity.

Figure 3 shows the distribution of message samples for the seven specific situations and Table 4 shows the frequency of topics for each situation. The frequency reported for the kind of text messages and their associated situations is elicited from the initial samples provided by participants as well as samples obtained from the second step (augmentation process) that shared the same contextual information. For instance, if 5% of message samples that are provided by participants while they are nearby a point of interest concerned asking for help to get to a location, for example 'I am here, come out and find me' or 'I am here and need help', then we added messages retrieved from the Weibo dataset (10%) that shared the same contextual information. We summed all the samples associated with the 'outdoor point-of-interest' situation provided by participants and those retrieved from the Weibo dataset and reported the percentage (15%) over all the message samples (850+) in our dataset. In the remainder of this section, we explain the most occurring message instances for each specific situation.

Out of 850+ of our total message samples, 30% are most relevant to in-vehicle situations. When users are in-vehicle they often inform others that they are on their way. In this situation, they might add other details to their messages, such as asking for help to get to a location once they arrive or offering help to others by asking them if they would like to share the ride. Users also send messages to notify others that they will arrive late. In this situation, they might coordinate the arrival time, reschedule existing plans, and request or offer a ride. Another frequent type of text message users send is to inform or ask others about the arrival time. When users specify their arrival time, they might ask others to pick them up at the destination or guide them to the location. Additionally, they might cancel or reschedule their plans based on the estimated arrival time.

Overall, 23% of message samples are relevant to waiting at stations and 11% are related to walking to stations. When users are waiting at the bus/train station, most text samples concern notifying others that they are waiting.

Table 3. Examples of message samples for 'arrive late' and 'bring food' topics along with alternative values for selected parametric message instances.

Reference sample	Messages	Values	Sub Values
I will **arrive late** (entailment)	I will be an hour and a half late	My son, mom, my friend, dad, my daughter, my wife, my husband, my colleague, the speaker, our guest – 5 min, 10 min, 15 min, 20 min, 25 min, 30 min, 35 min, 40 min, 45 min, 50 min, 55 min, 1 hour	
	I came back late!	My son, mom, my friend, dad, my daughter, my wife, my husband, my colleague, the speaker, our guest	
	. . .		
	The plane is late!	Train, bus, shuttle	
	I will be an hour late for dinner!	My son, mom, my friend, dad, my daughter, my wife, my husband, my colleague, the speaker, our guest – 5 min, 10 min, 15 min, 20 min, 25 min, 30 min, 35 min, 40 min, 45 min, 50 min, 55 min, 1 hour – lunch, breakfast, the appointment, the meeting, the talk, the interview, meal, the class, the session	
Could you **bring food?** (entailment)	Does anyone bring food for me?	My son, mom, my friend, dad, my daughter, my wife, my husband, my colleague, the speaker, our guest	
	Can you bring breakfast at 2?	7, 8, 9, 10	05, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55
	Could you bring food for breakfast?	Food, tea, coffee, biscuits, bacon, pancakes,cookies, milk, bread	
	. . .		
	I am going to bring pizza for lunch	Food, sandwich, cookie, soup, bread, coffee	
	Let's bring snacks for the picnic	drinks, coffee, tea, sandwiches	
Could you **bring food?** (contradiction)	Sorry, I can't brink meal today	my colleague, my friend, my family, my children, my husband, my wife, my children – food, breakfast, lunch, dinner – tomorrow	
	I can't bring breakfast on Monday	my colleague, my friend, my family, my children, my husband, my wife, my children – food, meal, lunch, dinner – Tuesday, Wednesday, Friday, Saturday, Sunday, weekend	
	I already had lunch, no need to bring food	my colleague, my friend, my family, my children, my husband, my wife, my children – food, breakfast, dinner	
	Today I will skip lunch, no need to bring food	Tomorrow – my colleague, my friend, my family, my children, my husband, my wife, my children – food, breakfast, dinner	
	. . .		
	I can't bring lunch at 1, can I bring at 3 instead?	12, 1, 2, 3, 4 – 12, 1, 2, 3, 4	05, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55 – 05, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55
	Mom can't bring dinner today	my colleague, my friend, my family, my children, my husband, my wife, my children – food, meal, lunch, breakfast – tomorrow	

In some instances, they might specify the amount of time they have been waiting. If users have been waiting for a long time, they might ask others for a ride or cancel existing plans. Additionally, while waiting at the station, they might send a reminder to others about a location where they or others have to wait to meet each other. Users also set up phone calls or send reminders about the time they planned to make a phone call. Alternatively,

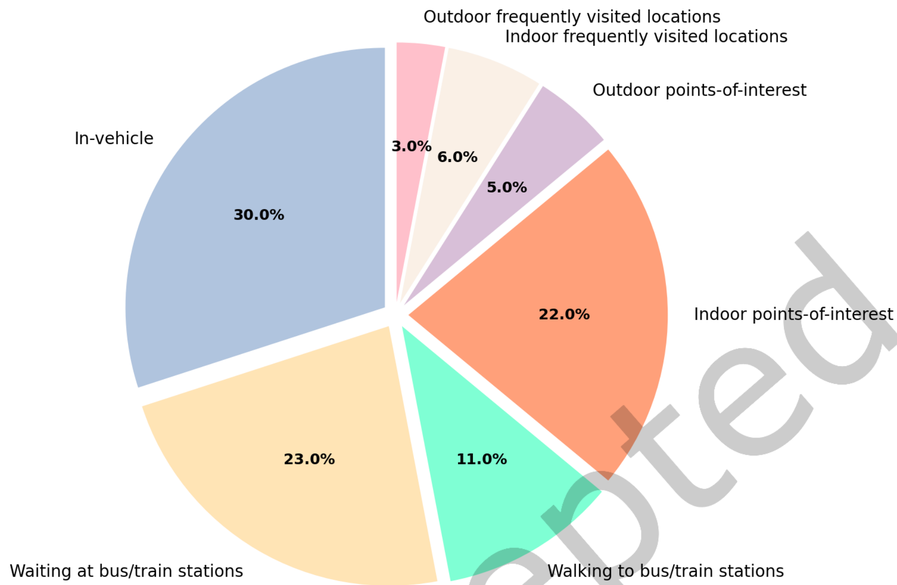


Fig. 3. The distribution of the percentage of the 850+ messages in our sample across the seven "on-the-go" situations in which the BVI participants in our prior study [35] indicated they typically send a text.

they might schedule/reschedule appointments and meetings. When users are walking to reach the bus/train station, most instances of messages inform others that they might arrive late. Users might also remind others that they are going shopping, asking others for a ride to go to a supermarket, and making plans with others for shopping. Other frequent messages notify others that they are on their way, ask others to be picked up from the station by specifying the arrival time, and communicate about appointments.

Out of all the messages, 22% are relevant to outdoor points-of-interests (POIs) and 5% are related to indoor points-of-interests. When users are outdoors, most message instances concern requesting to be picked up or notifying others that they are on their way to pick them up. Users also ask others for help getting to the location, especially when they cannot find the entrance. Other frequent messages concern signaling the location as well as informing others that they are waiting. When users are indoors, most messages are relevant to inform others about the time they arrived at the destination and request help to get to a location. Other frequent messages might concern checking in the location of the person they planned to meet and letting others know that they might arrive late.

Finally, 6% of our message samples are relevant to indoor frequently visited locations (FVLs), and 3% are related to outdoor frequently visited locations. When users are indoors, most occurring text messages request others to bring food. On other occasions, users send a text about a meeting, such as the starting time, letting others know that they are in a meeting, and reminding others about an upcoming meeting. In addition, users might

Table 4. Frequency of text messages associated to each topic for seven specific situations. Shaded colors are the most occurring topics in each situation.

Topic	**In-vehicle**	**Waiting at stations**	**Walking to stations**	**Indoor POI**	**Outdoor POI**	**Indoor FVL**	**Outdoor FVL**
On my way	17.97%	2.05%	8.51%	2.27%	0.53%	0%	4%
Arrive late	16.81%	6.63%	14.89%	9.09%	1.07%	3.85%	0%
Arrival time	16.40%	4.08%	7.45%	43.18%	0%	3.85%	0%
Pick up	7.82%	5.10%	8.51%	2.27%	22.04%	1.92%	0%
Waiting	6.64%	11.23%	0%	4.55%	11.83%	0%	12%
Get to a location	0%	0%	2.13%	11.36%	17.74%	0%	12%
Shopping	7.03%	6.12%	14.89%	2.28%	5.38%	3.85%	0%
Signal my location	4.69%	5.10%	1.06%	6.82%	10.22%	1.92%	12%
Appointment	3.12%	9.69%	8.51%	2.27%	4.3%	5.77%	4%
Meeting	3.12%	9.18%	7.45%	6.82%	2.69%	13.46%	4%
Phone call	1.56%	10.71%	1.06%	0%	6.99%	11.54%	0%
No able to talk	0.39%	3.57%	1.06%	0%	8.07%	13.46%	0%
Bring food	7.03%	1.02%	4.26%	0%	3.76%	36.54%	36%
Check-in location	2.74%	6.12%	5.32%	9.09%	5.38%	3.84%	12%
Check-in guest	1.56%	7.65%	0%	0%	0%	0%	0%
Making plans	1.56%	8.17%	4.26%	0%	0%	0%	0%
Changing plans	1.56%	3.58%	10.64%	0%	0%	0%	4%

let others know that they are not able to talk or schedule a phone call. Similarly, when users are standing at frequently-visited outdoor locations, they often communicate about bringing food. Users also send text messages to signal their location by providing information about nearby landmarks, check-in the location of the person, letting others know they are waiting, and requesting help to get to a location.

Overall, including sentences without parametric message instances, we were able to generate over 850 message samples that can be retrieved and filtered by the topic of conversation (e.g., arrive late, get to a location) and user's situation (e.g., in-vehicle, nearby POI). These samples are distributed over 17 specific topics and range from 19 message samples for the 'check-in guest' topic to 81 samples for the 'pick up' topic. By adding the associated values to message samples, we were able to expand the set of message instances to over 250k messages. Associated values include alternative phrases that can be used to personalize a message. For example, if a sample is "my Uber is late", users can personalize the message by replacing 'Uber' with other values such as 'bus' or 'train'. We have made our dataset publicly available: https://github.com/Banus/Textflow/blob/main/Dataset.tsv

5 TEXTFLOW: MIXED-INITIATIVE AND SCREENLESS TEXTING

Based on the results of our study and the initial message samples obtained from our previous work [35], we developed TextFlow (https://github.com/Banus/Textflow), a mixed-initiative text messaging system that enables BVI users to receive and interact with suggested messages relevant to their immediate situation, while at the same time bypassing the reliance on a reference visual screen. The system is made of three parts: a User Model, a Reasoning Model, and a Task Model (Figure 4).

5.1 The User Model

The User Model captures key aspects of the user's daily routine. The model adapts to the specific needs of BVI users for manipulating a text message and contains four types of data: the user's context, the user's situation, personal-related commitments, and a dynamic profile. To formally define the elements of the User Model, inspired by [13], the user's context C is defined as the ontology of location, time, and activity and we represent it as

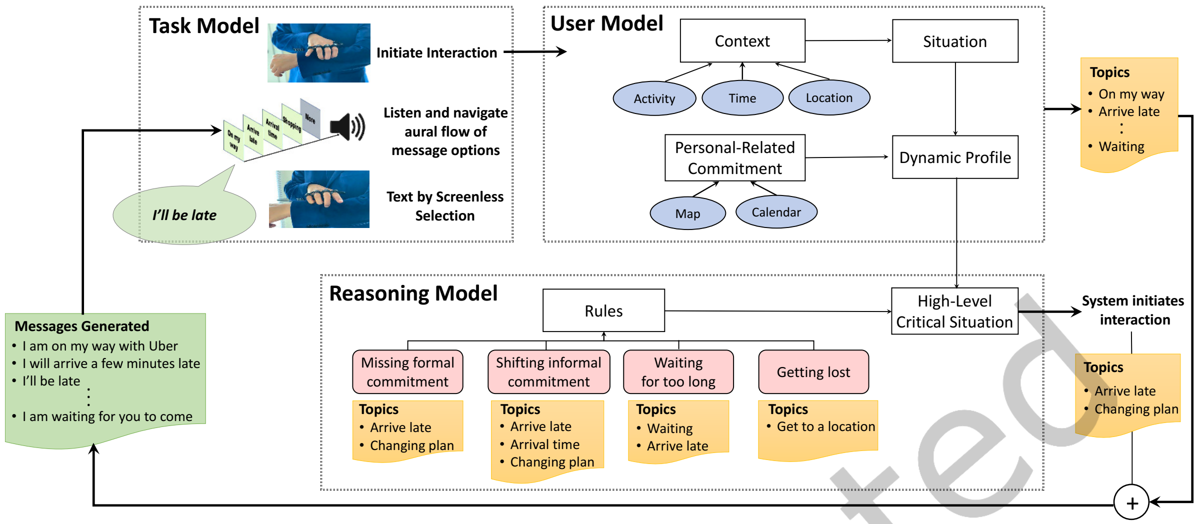


Fig. 4. Three parts compose TextFlow: 1) the User Model leverages user's situational factors to generate a dynamic user profile; 2) the Reasoning Model operates on four rules based on the topics elicited from the initial study; 3) To support the user interaction, the Task Model enables the blind to listen to and select aural prompts in screenless mode via an existing finger-worn device (TapStrap).

$C = (O_{Location}, O_{Time}, O_{Activity})$. An instance of the user's context is a user's situation, defined as:

$$S = (O_{Location}\ x_i, O_{Time}\ x_j, O_{Activity}\ x_k)$$

where x_i is the location instance, x_j is the time instance, and x_k is the activity instance. We define two types of locations, retrievable from Google Maps: known locations and unknown locations. Known locations require either the user to manually label work, home, and school locations or the system to recognize a point of interest (e.g., train stations and restaurants). Locations that are not recognized as known are categorized as unknown locations. The location is represented as a longitude, latitude pair (Log, Lat).

To detect points of interest and labeled locations, we set a threshold to 0.1 miles on the distance from the user's current location. Time is divided into three periods of morning (5am-12pm), afternoon (12pm-5pm), and evening (5pm-9pm). The user activity is classified into three categories of still, walking, and in-vehicle. We used an activity detection model based on sensor data to classify the user's activity automatically. A confidence score is assigned to each activity and the one with the maximum score is detected as the user's current activity.

Personal-related commitments (PRCs) signal the user's recurrent and non-recurrent routines, and they are elicited from the personal calendar available on the mobile device and the data from the map. The calendar includes a list of commitments for each day with title, location, start and end time. The location must belong to one of the known locations and its coordinates are retrieved from the map. By combining the user's PRCs and the current situation we generate a **dynamic profile** that integrates real-time data from the user.

When the user initiates an interaction for sending a message, the system suggests topics and messages based on the User Model data. Topics are the entry points of TextFlow, and their order is determined by the user's situation, which is constituted by the type of activity, location, and time (see 1). When the activity is in-vehicle the system re-ranks topics for an in-vehicle situation regardless of the location. When the activity is classified as still or walking and the location is unknown, the system re-ranks topics for an on-foot situation. If the location is recognized as known, the system re-ranks topics either by frequent location or point of interest. Each topic corresponds to a list of messages, such that some of the messages depend on the location, time of the day, and PRC (see Table 2).

5.2 The Reasoning Model

The Reasoning Model provides rules to detect inconsistencies and infer new data [45]. These rules lead to 'high-level critical situations' [13], a class of situations in which the system initiates the interaction and selects topics bound to a specific rule. Based on the situations and messages learned from the initial study, we define four generic rules to determine when the system initiates the interaction and what topics are presented to the user.

Missing formal commitments: BVIs may miss important commitments due to limited access to time and traffic information. For each formal commitment, we estimate the arrival time based on the speed of the user. While users are en route, starting 15 minutes before their scheduled commitment, the distance to the destination is recorded every 5 minutes and the average speed is computed as the distance covered in the last interval. The arrival time is estimated by extrapolating the average speed to the destination, and the rule is activated when the arrival time exceeds the time of commitment by 10 minutes. An example is when a person should be at work before 9 am, and the current situation is $S = (in - vehicle, 8{:}45am, \forall location)$. If the user's distance from work is 26 miles and 5 min later is 21 miles, the rule will be activated. The selected topics for this rule are 'arrive late' and 'changing plan'.

Shifting informal commitments: When BVIs are busy with one activity, they may forget to inform others about informal commitments, such as meeting a friend at Starbucks or going home. Shifting informal commitments are elicited from PRC based on the ending time of their formal commitments. In this study, the rule is activated when a person is not in the expected location or en route 15 minutes after the end of the commitment. An example is when a person leaves their place of work at 5:30pm and the situation is $S = (still, 5{:}45pm, work)$. Chosen topics for this rule are 'arrive late', 'arrival time', and 'changing plan'.

Waiting for too long: When BVIs wait in a specific place they may not realize the length of the time they spent there. Waiting for too long is defined as when the user stays still for at least 30 minutes and the location is labeled as 'station' or 'unknown'. An example is when a person is in the bus station and the situation is $S = (still, 2{:}15pm, bus - station)$ and after 30 minutes it is $S = (still, 2{:}45pm, bus - station)$. The selected topics for this rule are 'waiting' and 'arrive late'.

Getting lost: Finding the exact location of known places is not always feasible. While GPS is a possible solution, it may not always be accurate enough to support BVI users. In our study, getting lost occurs when a person is walking for at least 15 minutes in the same general area outdoors. An example is when a person has an appointment with their friend in a restaurant and after reaching the place the situation is: $S = (walking, 2pm, restaurant)$ and after 15 minutes is: $S = (walking, 2{:}15pm, restaurant)$. The selected topic for this rule is 'get to a location'.

5.3 The Task Model

Task models incorporate the logical steps associated with activities that an individual must perform to achieve their goal [23]. With TextFlow the user is either notified proactively of relevant topics based on the situation or the user can initiate the texting action, and this input triggers the User Model. A list of relevant topics and messages are offered to the user for interaction and topics are ordered based on the frequencies obtained from the initial user study. Topics with high frequency (shown in Figure 1) are read individually, whereas topics with a frequency of 2 or less are collected in a 'More' category, which plays on-demand after the last topic (Figure 5-c). These outputs are then fed into the Task Model.

The interaction is supported by a wearable device called TapStrap [29] (Figure 5-b) which uses finger movements to perform various operations such as selection or deletion to send a message via a messaging app (Figure 5-a).

The system supports two different interaction modalities: self-disclosing flow (SDF) and topic-by-topic (TBT) browsing. Topic-by-topic browsing allows the user to navigate the system's suggestions with explicit forward and back commands, whereas self-disclosing flow reads the system's suggestions sequentially. We iteratively engaged five sighted users in exploring different combinations of finger commands to control the flow of topics and messages, and over 4 weeks, we converged on a set of commands that were supported and reliable by the device and at the same time used easy-to-activate fingers. We then mapped each finger to an operation upon agreement between users (see Figure 5-b). We describe the steps that the user has to perform when interacting with each of the modalities:

Topic-by-topic browsing: In this modality, users have the ability to scan the system's suggestions one-by-one, and this allows them to control the speed of browsing topics/messages. The main rationale for the design of this modality is based on the fact that screen reader users grew accustomed to have a fine-grain control over structured lists of items rendered in a text to speech format: they are used to scan item by item links on a webpage, content headers, and navigation functions [43]. So, in this case, the TBT interaction modality leverages screen reader users' familiarity with the ability to have more precise control over auditory lists of options. For this interaction modality, users tap their index finger to listen to topics individually. They can navigate backward through suggestions by using the middle finger, tap their thumb to select a topic and then browse the message list in the same way. Users have the option to tap their ring finger to cancel the selection. When the system initiates the interaction, users can either tap their pinky finger to ignore the system's suggestions or continue browsing topics one by one.

Self-disclosing flow: To mitigate the need for using multiple fingers, we introduce an alternative TextFlow layout in which the system's suggestions are self-play, i.e., read out sequentially. We define three different dwell times, that is, the pause between the system suggestions. The first is node-to-node dwell time (150ms), which is the pause between topics or the pause between messages. The second is the dwell time between topics and messages (1s). The third is the flow-to-flow dwell time (1s), that is, upon hearing the intended topic, the user can tap their thumb to make a selection, and if no selection is made the system will repeat the topics from the start after the dwell time. The user can tap their index finger to cancel the selection. Upon selecting a topic, the system will transition to messages. Similarly, the user can tap their thumb or index finger to make or cancel a message selection respectively.

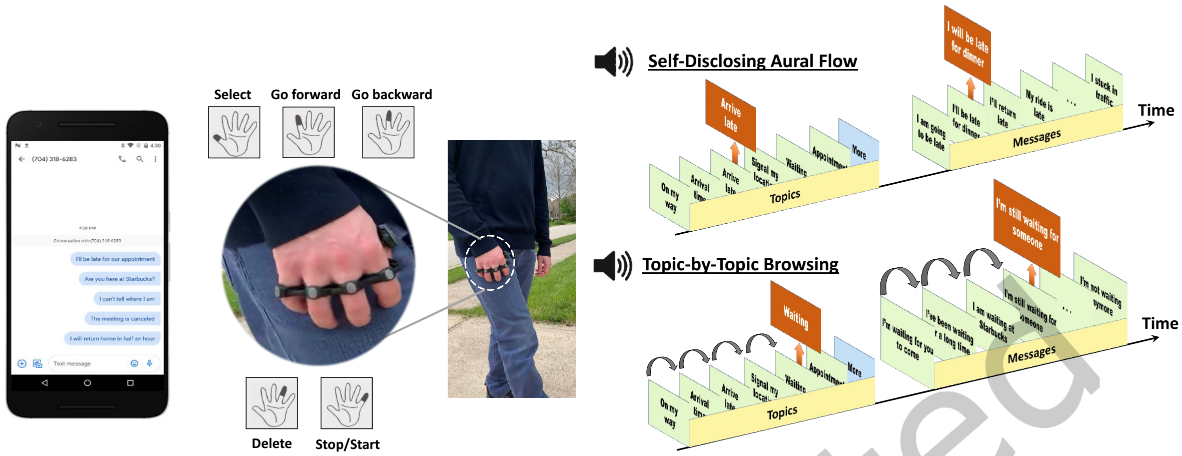
We implemented and deployed TextFlow on a Nexus 6x phone with Android 10.0 and Android SDK v.29.0; the TapStrap device was interfaced using Android Tap SDK 0.3.3, while the aural rendering of the message options was realized through the default Android text-to-speech (TTS) engine [20]. To feed contextual data, we used Google Location and Context services [22] for the points of interest, the Google Calendar API [21] for personal-related commitments (PRC), and the Android Activity Recognition API [19] for data on the current user activity.

6 STUDY WITH BLIND PARTICIPANTS

We conducted an evaluative, in-person user study (IRB-approved) of TextFlow to scrutinize the usability and the user experience with the system for first-time users. For the purpose of evaluation, we used the initial message samples obtained from our previous work [35].

6.1 Participants, Setting and Procedure

We recruited 10 BVI individuals (6 females and 4 males) in the Indianapolis urban area. Six of them already participated in the initial study (Section 3). The participants' ages ranged from 33 to 68 years old (μ=47.7, σ=10.37). Eight used iOS and two used Android. The length of mobile usage ranged from 6 to 18 years (μ=11.8, σ=4.66). Five participants were identified as totally blind, three legally blind, and two had retinopathy of prematurity (ROP).



(a) (b) (c)

Fig. 5. Detailed view of the Task Model. Users browse topics and messages either as self-disclosing aural flow or by scanning them one-by-one (c); Topics and messages are selected through an off-the-shelf finger-worn device (TapStrap[29]) that supports five tapping gestures (b) and sent via a messaging app on the mobile device (a).

The study was conducted in a conference room, with a table and chairs. Each participant remained standing while interacting with our system and could rest, sit or take a break at any time. Based on our first study, we defined three scenarios to capture specific situations where a blind person might need to send a message: (1) On my way to meet a friend at Starbucks at 2 pm; (2) On my way to work riding with a friend; (3) Leaving the house at 1 pm to meet a friend for lunch. For each scenario, we asked participants to complete five tasks, such as "Letting a friend know I'll be late", or "Reminding a colleague about an upcoming meeting". Our study utilized a within-subject design where each participant completed 15 different tasks distributed evenly over three scenarios. In the first and third scenarios, four tasks required the user to initiate the interaction with TextFlow, whereas one task was associated with a rule in which the system initiated text suggestions. In the second scenario, two tasks were associated with rules and three tasks required the user to start messaging. When the user started messaging, the default interaction modality was topic-by-topic browsing. However, when the system initiated the interaction, the default modality was self-disclosing flow. The order of the three scenarios was counterbalanced to account for any ordering effects. Each scenario was executed indoors and described verbally to each participant to help them understand both the situation and the type of message they could send for each task. Participants were first introduced to the TapStrap, the hand position required while tapping with the device, and the role of each finger in sending a text message. In the first phase, we explained the purpose of each topic with one or two examples of text messages associated with the topic and introduced the two interaction modalities—topic-by-topic browsing and self-disclosing flow. In the second phase, participants were asked to practice and send a text message using the two modalities. Participants were provided with feedback and assistance if needed during the second training phase.

After training, participants were required to start the tasks associated with the first scenario, and each scenario was read aloud by the researcher. During the interaction, users could tap on any part of the body that was convenient for them. Once the user sent a message, they were asked to determine whether the selected message was relevant to the context or not. Each scenario was video recorded, followed by a five-minute break. After the last session, we asked participants interview questions. The questions focused on situations in which the system could be useful, advantages and disadvantages of the system versus accessible keyboards and dictation, and comparison between the two interaction modalities. Answers to interview questions were audio recorded. For approximately two hours of participation, each participant received a $60 Amazon gift card.

6.2 Results

6.2.1 ***User Performance analysis.*** To assess efficiency, we tracked the time participants took to send a message and computed the average time (μ) and standard deviation (σ) for task completion across all participants. To assess effectiveness, we defined and analyzed key measures of the task success for each task and participant. Furthermore, to compare the two interaction modalities (topic-by-topic browsing and self-disclosing flow) we computed interaction error and error rate. Finally, to assess the learnability of the system, we computed the average cancelation rate and the average back rate over the three sequential scenarios.

Time on task: For the 11 tasks associated with topic-by-topic browsing, the time spent depended on where the target topic and message were in the auditory flow. Out of 11 tasks, six had both the target topic and the message among the top in the list and yielded the highest performance (μ= 17.42; σ= 9.43), between 10.5 to 17.5 seconds. For three out of 11 tasks, either the target topic or the message was close to the end of the list: it took longer to identify and select the message (μ = 33.97; σ = 11.3), between 29 to 35 seconds. Finally, for the two tasks where both the target topic and the message were at the end of the list, users selected the target message between 42 to 44.5 seconds (μ = 43.7; σ = 6.48). The narrow range of standard deviations across all tasks (between 5.58 and 13.48 seconds) indicates that the performance was quite uniform across participants, with few exceptions. The exceptions are due to two main factors: the number of times that the user had to go backward (18 times in 11 tasks) and the number of times they had to cancel the selected topic (15 times in 11 tasks). The former occurred when the user browsed system suggestions very rapidly, inadvertently skipping the target topic or message. The latter happened when the user could not find the intended message within the first topic and had to get back to the list to select another topic. For the 4 tasks associated with the self-disclosing flow, the average time (μ = 35.6) ranged between 28.7 to 46.3 seconds, while the standard deviation (σ = 23.52) sat between 18.30 to 27.92 seconds. Two factors affected the time spent on these tasks. The first factor was the number of times a selected topic or message was canceled (30 times in 4 tasks): when participants failed to select the target topic or message within one second, the system selected the next suggestion. The second factor was that 3 out of 10 participants preferred to listen to the system's suggestions for one or two loops to memorize choices and avoid canceling a selection.

Task success: For each task, we identified one or multiple target messages, defined as the messages that were designed and generated as most pertinent to a given situation. To determine task success, we defined the following criteria: Success (1): the participant selected the target message and confirmed it was relevant to the situation; Partial Success (0.5) captures two cases: when the message was off target but perceived as relevant; or when the message was on target, but the participant found it not relevant; Failure (0): the message was off target and deemed not relevant by the participant. To compute the task success, we normalized the value assigned to each task over the number of participants. The average task success rate across the three scenarios is 88.66% (see Figure 6). On average, 4 out of 15 tasks had a 100% success rate across all 10 participants. In the remaining 11 tasks, an average of 1.6 participants per task had partial success and an average of 0.72 failed. We observed that when users had to continuously search by either browsing over more topics or canceling their initial choices, they forgot the details of the task and ended up selecting a message that was either outside the target message or perceived as not relevant.

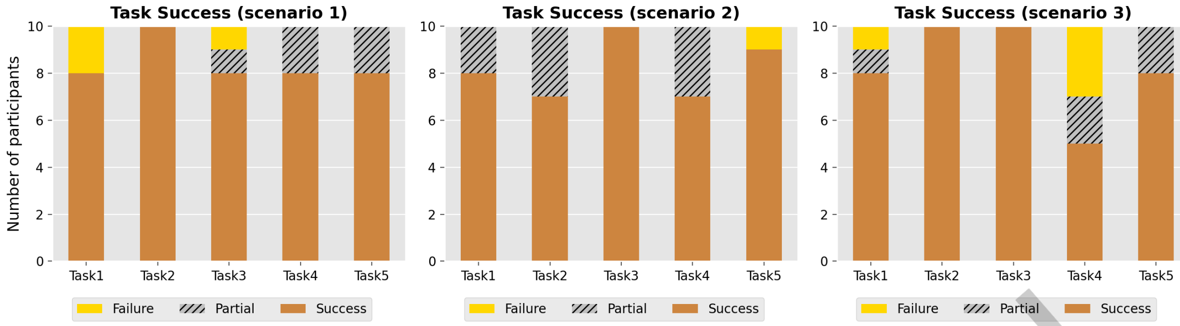


Fig. 6. All participants successfully sent screenless messages. In five tasks, two participants on average selected a message off-target; in nine tasks, two participants on average did not perceive the message as relevant to the situation.

Interaction Error: We defined errors as the number of times that users selected a topic or a message that was off target and led them to cancel the selection. We compared the number of errors for the two different interaction modalities—*topic-by-topic browsing* (TBT) and *self-disclosing flow* (SDF)—for each participant across all three scenarios. For the four tasks associated with the SDF condition, 8 out of 10 participants made errors (with an average of 0.77 error per participant) with a total of 31 cancellations and two participants had no error. For the 11 tasks associated with the TBT condition, 9 out of 10 participants made errors (with an average of 0.15 error per participant) for the total of 18 cancellations and one participant had no error. To compute the error fraction, we summed the number of errors for the two conditions for each participant and normalized the values (Figure 7). The result indicates that when the system was in the SDF condition, the error fraction was higher, whereas when the system was in the TBT condition, participants had more control over the auditory lists of topics and messages, which led to fewer errors. When conducting a one-tailed T-Test we found the difference in the number of cancellations between the two conditions to be significant ($p = 0.003 < 0.05$).

From our observation of the participants and interview analysis, when the system was in the SDF condition, interaction errors were mainly due to two factors: the short time between the auditory list of topics and messages and users' unfamiliarity with the message options in the first trials. When the system was in the SDF condition, users had only one second to select the target topic or message before the system read the next option in the auditory list. As a result, participants selected the wrong topic or message in their initial attempts. When asking users about their thoughts on the SDF interaction modality, some participants mentioned it will be beneficial if the system can put a sound (e.g., beep) between each option in the auditory list instead of a pause. In addition, users specified that the more they interact with the system, the better they can memorize the topics and messages, which in turn can reduce the number of errors they will encounter for the SDF interaction modality. When the system was in the TBT condition, a few participants browsed topics or messages very rapidly, which resulted in selecting the wrong option. Similarly, participants commented that interacting more with the system could help them to learn the options and to better control the speed of browsing. For example, when users knew that the message option is among the last in the auditory list, they could browse faster in the beginning to pass the initial messages.

Learnability: We operationalized learnability as the number of times participants canceled a selection and/or went backward when they skipped a message or a topic over the three sequential scenarios. When the system was in the TBT condition, cancelation rate and back rate for each participant were calculated by dividing the number of times users canceled their selection and went backward by the number of tasks in each scenario,

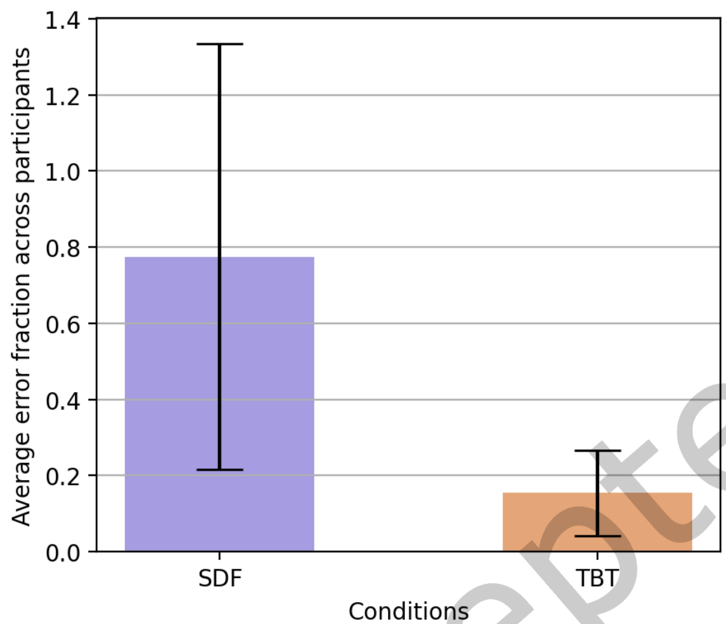


Fig. 7. On average, participants made more errors when the system was in the SDF condition compare to when the system was in the TBT condition.

respectively. Similarly, when the system was in the SDF condition, we computed the cancelation rate by dividing the number of times users canceled the selected topic or message by the number of tasks in each scenario. Figure 8 shows the average cancelation rate and the average back rate for the TBT condition and the average cancelation rate for the SDF condition among all participants over the three scenarios. When the system was in the TBT condition, 7 out of 10 participants made fewer errors as they moved from scenarios 1 and 2 to scenario 3, two participants (P7 and P8) had no errors across all three scenarios, and one participant (P3) made more errors in the third scenario. In addition, 5 out of 10 participants went backward fewer times while browsing the system's suggestions as they moved from scenarios 1 and 2 to scenario 3. Two participants (P4 and P6) went backward further as they moved from scenarios 1 and 2 to scenario 3. Two participants (P8 and P10) did not go backward across all three scenarios, and one participant (P9) went backward the same amount of time in scenario 1 and scenario 3 but did not move backward in scenario 2. When the system was in the SDF condition, 7 out of 10 participants made fewer errors as they moved from scenarios 1 and 2 to scenario 3. Two participants (P4 and P9) kept a level of zero error across all three scenarios. One participant, P2, made the same number of errors in scenario 2 and scenario 3 but no error in scenario 1. These results indicate that as participants interact more with the system the learnability improves, allowing them to select the target topic or message in the first attempt. Participants commented that by moving from scenarios 1 and 2 to scenario 3 they were able to memorize the topics and messages, resulting in fewer instances of canceling the selection or going backward.

Error rate: To better understand the relationship between interaction errors and learnability, we computed the error rate as the percentage of participants who canceled their selection one or more times for each task across the three scenarios. When the system was in the TBT condition, the average error rate was 13% for the four tasks

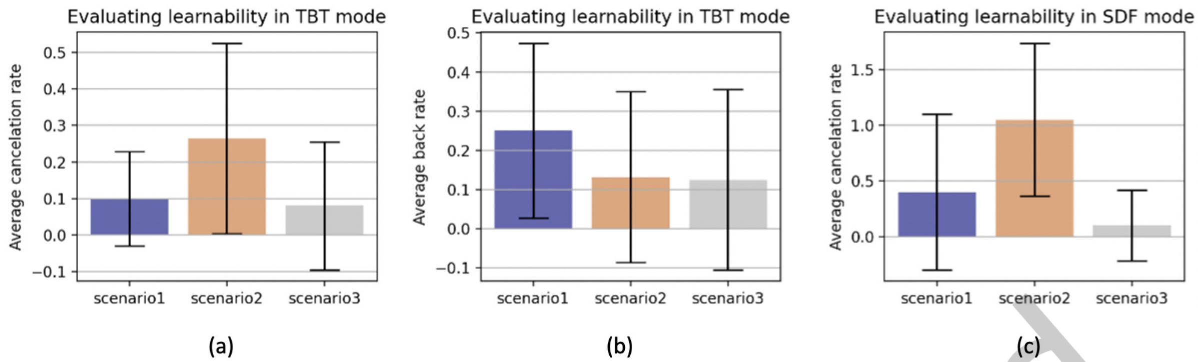


Fig. 8. Average cancelation rate for all participants decreased when moving from scenario 1 and 2 to scenario 3 in both the TBT and SDF conditions (a) and (c); Average back rate decreased for all participants when moving from scenario 1 and 2 to scenario 3 (b).

in scenario 1, 23% for the three tasks in scenario 2, and 8% for the four tasks in scenario 3. On the other hand, when the system was in the SDF condition, the average error rate changed from 30% for one task in scenario 1, to 60% for the two tasks in scenario 2, to 10% for one task in scenario 3 (see Figure 9). The results indicate that when the system was in the TBT condition, fewer participants made errors compared to when the system was in the SDF condition. This confirms the results reported for the error interaction, which stated when users have control over the auditory lists of options and can browse the system's suggestions one by one, the usability is better compared to when they need to listen to auditory lists of topics or messages and make a selection. In addition, the error rate for each participant decreased in scenario 3 for both the TBT and SDF conditions. This confirms the reported result that as participants interact more with the Textflow, the learnability of the system improves.

6.2.2 ***Interaction Behavior Analysis.*** To analyze in more detail the interaction behavior with TextFlow, we coded the video-recordings of each participant to capture the following constructs: (1) *fluidity*, (2) *difficulty*, (3) *hesitation*, (4) *finger position*, and (5) *hand placement*.

Fluidity: We defined fluidity as interaction instances in which users perform a selection accurately, rapidly, and with no assistance.

Difficulty: When users fail to perform an operation smoothly but require more than one attempt, we coded that event as difficulty.

Hesitation: Hesitation refers to when a user's finger movements remain idle for more than two seconds before performing an operation.

Finger position: Finger position refers to the way that users hold their fingers to interact with the device (e.g., flat and clench)

Hand placement: Hand placement refers to the place on the user's body where they choose to put their hand (e.g., leg and arm).

The video of each participant is divided into 5 segments per scenario, such that the start and end of each segment refer to users' interactions associated with one task. Two individuals from our research group coded 10 videos each after discussing and agreeing to the codebook.

To compare the interaction behavior across the two conditions, *topic-by-topic browsing* (TBT) and *self-disclosing flow* (SDF), we coded the 11 tasks associated with the TBT condition and the 4 tasks associated with the SDF condition. We then counted the tasks that are coded as fluidity, difficulty, or hesitation in each condition for

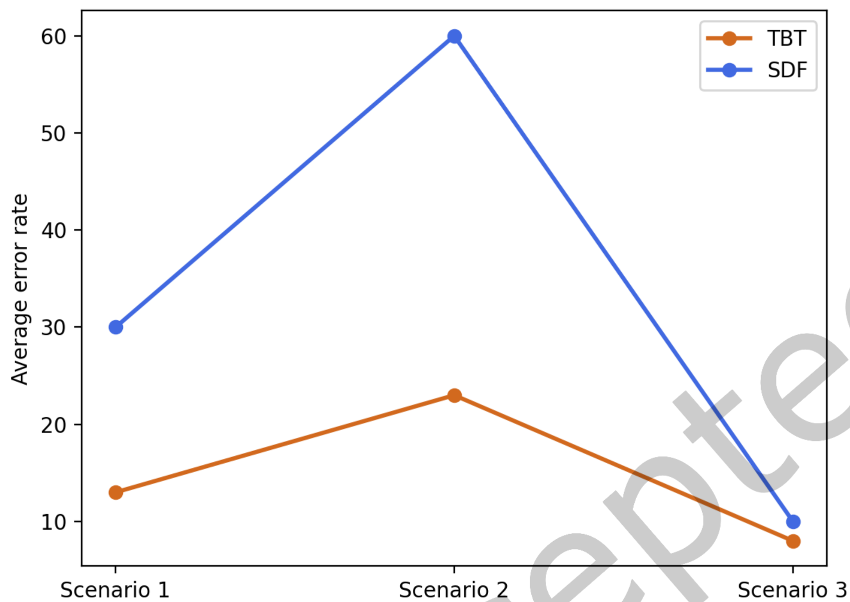


Fig. 9. Average error rate among all participants decreased when moving from scenario 1 and 2 to scenario 3 with the higher average error rate for each scenario when the system was in the SDF condition.

each participant. In addition, we counted the number of tasks coded as fluidity, difficulty, or hesitation for three different combinations of hand placement and finger position. Note that the last two constructs (finger position and hand placement) depend on the assessments of the first three.

Overall, our results demonstrate that most participants were able to send a text message accurately, rapidly, and with no assistance. When the system was in the TBT condition, on average, 9.09% of participants required more than one attempt to send a text message. Based on our observation, errors were caused by two main factors: (1) rapid browsing of topics and/or messages, (2) semantic similarity of some topics. When participants browsed the system's suggestions rapidly, they could make a wrong selection and had to cancel and scan the system's suggestions one more time. Based on the interview analysis, participants preferred to personalize some of the topics or messages, such that they can access the most frequently used topics/messages among the first options in the auditory list to avoid browsing many options. We believe adding a personalization component will reduce cognitive load and help users to select the target topic/message in the initial attempt. Additionally, when some topics had semantic similarities, such as 'pick up' and 'bring food', participants chose one, but they were not able to find the intended message associated with the selected topic and thus had to cancel the selection. For instance, when the task asked to send a message to a friend to bring lunch in the afternoon, a few participants chose the topic 'pick up' instead of 'bring food' and thus were not able to find the appropriate message among the options. Participants expressed the desire to interact with the system over a few weeks to learn and memorize the options in the auditory list. This indicates that including a training session, in which participants can become familiar

Table 5. Percentage of tasks coded as fluidity, difficulty, and hesitation in TBT and SDF conditions across all participants.

	Fluidity	Difficulty	Hesitation
TBT	80.0%	9.09%	10.91%
SDF	45.0%	40%	15%

with the list of topics and messages could potentially reduce the number of errors due to semantic similarity. On average, 10.91% of the times users' fingers remained idle for 2 seconds or more before making a selection. When participants were browsing messages associated with a topic, they spent some time thinking before making a selection. This situation happened when a task had two or more target messages that participants could select from. Similarly, in this situation adding a training session could help to alleviate the confusion due to multiple target messages associated with a task.

When the system was in the SDF condition, fewer participants interacted fluidly with the system compared to the TBT condition, due to fast-spoken topics and messages. In most instances, users ended up selecting the next option after hearing the target topic or message, which led to multiple attempts (40% difficulty). Alternatively, other participants preferred to first memorize the options and their order in the auditory list and then make a selection. As a result, their fingers remained idle for 10 seconds or more (15% hesitation). Based on the interview analysis, when the system was in the SDF condition, users preferred to primarily learn the options and their order in the auditory list to avoid making a wrong selection. Table 5 shows the average number of tasks for the three interaction behavior codes for each condition across all the participants.

Our results show that most participants (80%) interacted with the Textflow by placing their hand on their leg and 20% of participants placed their hand on their chest. Half of the participants who placed their hand on their leg decided to clench their fingers while the other half kept their fingers flat. All participants who placed their hands on their chest decided to keep their fingers flat while interacting with the system. When the system was in the TBT condition, 4 out of 8 participants who had 'leg' as hand placement and 'clench' as finger position encountered difficulties sending a text in executing 6 tasks and hesitation in executing 5 tasks. When the system was in the SDF condition and participants had the same hand placement and finger position, the number of tasks coded as 'difficulty' was 8, whereas the number of tasks that were coded as hesitation was zero. Participants who had 'leg' as hand placement but kept their finger position 'flat', encountered difficulties sending a text one time for the TBT condition and 5 times for the SDF condition. Moreover, 6 tasks were coded as hesitation both in the SDF and TBT conditions. Finally, for the two participants who chose 'chest' as hand placement with the finger position 'flat', 3 tasks were coded as difficulty and one task as hesitation in the TBT condition, whereas 3 tasks were coded as difficulty and zero tasks as hesitation in the SDF condition. Table 6 summarizes the results for the three combinations of hand placement and finger position among participants. As the results show, regardless of hand placement and finger position the number of tasks coded as fluidity was higher when the system was in the TBT condition versus when it was in the SDF condition. However, when the finger position was flat the number of tasks coded as hesitation was higher in the TBT condition, demonstrating that in this situation users paused more when tapping with their index finger versus when they listened to the auditory list of items and used their thumb to make a selection.

6.2.3 *Analysis of the Experiential Themes.*

The open coding thematic analysis of each participant's responses revealed that participants had a very positive experience with TextFlow and offered important insights, presented below:

Bypassing the phone: A theme that emerged from the participants' reflection on their experience is that TextFlow significantly reduces interaction barriers to access everyday texting in entirely auditory mode, especially when traveling. For example, P1 articulates how the phone-free approach of TextFlow can help him by saying:

Table 6. Percentage of tasks coded as fluidity, difficulty, and hesitation for three combination of hand placement and finger position in TBT and SDF conditions across all participants.

	Hand placement: **leg** Finger position: **clench** Num. of participants: **4** 			Hand placement: **leg** Finger position: **flat** Num. of participants: **4** 			Hand placement: **chest** Finger position: **flat** Num. of participants: **2** 		
	Fluidity	**Difficulty**	**Hesitation**	**Fluidity**	**Difficulty**	**Hesitation**	**Fluidity**	**Difficulty**	**Hesitation**
TBT	75%	13.64%	11.36%	84.09%	2.27%	13.64%	81.81%	13.64%	4.55%
SDF	50%	50%	0%	31.25%	31.25%	37.5%	62.5%	37.5%	0%

"*The system helps you send text messages without using your hands for the phone. And so, you could use your cane and fully concentrate on traveling as opposed to texting and traveling at the same time.*" Another participant, P6, remarked: "*It can be very useful when traveling and you have a whole lot of items with you so that you do not have to remove your cell phone and run the risk of losing it or having it stolen.*" P2 expressed a similar sentiment of comfort when able to rapidly access recurrent and ready-to-use messages, saying: "*I feel secure knowing that I got some quick access that I can contact my friend[s], to be able to say I am here, I am lost, or send my location.*"

When reflecting on daily situations, participants appreciated how by initiating and suggesting message options relevant to the context, TextFlow partially relieves blind users from having to constantly monitor their surroundings. P7 described an exemplary scenario, stating: "*If you're on a bus, you're trying to listen to when it's moving, or when it stops and what's the next stop. So, if the system could pull one or two topics up automatically based on the situation, then you don't have to be worried about the time or what is going on.*"

Privacy, efficiency, and accuracy: When reflecting on how TextFlow compares to dictation, users pointed to privacy, efficiency, and accuracy as key advantages with TextFlow. P4 stated: "*With dictation, I have to speak…everybody else can hear what I'm saying. But, if I was using this system and I had earphones on, nobody else can hear. So, there's privacy, which is a big deal for blind people.*" "*You might end up saying something two or three times, which makes Siri less efficient. Whereas with this, you can get to your message in a much shorter time.*"[P9] When comparing their use of TextFlow to accessible keyboards (e.g., VoiceOver), P6 shared: "*The thing about the keyboard is that a lot of times when you select the wrong letter, you have to delete and go back, whereas with this it's not like you're doing individual letters. You may send the wrong message, but you can go back and fix that.*"

Control preferred over self-disclosing flow: Participants expressed that browsing the messages one by one gives them more time to think before moving to the next item, whereas the self-disclosing flow requires heightened attention and more practice. P6 commented: "*when you browse, you can control the way it repeats them for you. The one that reads topics it takes more time before it becomes visible for most people and you have to memorize them first.*" Similarly, P3 stated: "*the browsing work[s] better for me because you can pick the speed at which you're going.*" Other participants indicated that the self-disclosing flow has the potential to better support multi-tasking. P2 stated: "*If I'm in the middle of doing something between the two choices I think the one that reads* *suggestions is better. I think that would be quicker. The browsing is better when you're calm, and not rushed.*" "*If you're walking using a cane, it'll be more convenient to use the one that the system reads the suggestions to you and you have to use only your thumb.*" [P1]

Envisioned use in additional scenarios: When reflecting on additional circumstances in which TextFlow could be useful, participants mentioned: 'finding a place', 'standing on public transportation', 'being in crowded places', and 'being in a meeting'. P7 commented: '*Even though I can get to the building, but I have a hard time finding a way to get in or to actually even get back out of the place. Because GPS works to get me to the spot, but it won't tell me where the door itself is.*""*if I got lost on the wrong street, I could text somebody to help me out faster, get me in the right direction.*" [P8] Another setting mentioned is public transportation, when it is difficult to type. P3 stated: "*It is useful like if you're standing in a train station or on the bus where your hands are full, but you could use the one hand to send a quick message whereas you wouldn't be able to type.*" P10 expressed that the system would be useful when standing in crowded places, stating: "*when I'm in a crowd of people, noise and like say, I'm standing at the door of the market.*" Participants also commented on the benefit of the system when they are in a meeting: "*When I am in a meeting, I can inform my family or friends without leaving the meeting or interrupting others.*" [P2]

7 DISCUSSION AND FUTURE WORK

Users who are blind or visually impaired face an additional overhead of mechanical constraints and cognitive effort to access basic mobile services such as texting. The problem has two sources. On the one hand, users have to overcome a screen-centric visual paradigm to interact with entirely aural prompts; on the other hand, current text entry methods require users to initiate text composition from scratch even for frequently sent messages. To address this problem, TextFlow uniquely combines in one solution three interactive approaches: (1) AI-driven, mixed-initiative interfaces that suggest situationally relevant content for candidate messages; the specific intelligent components of our system include the recognition of the user's activity (e.g., in-vehicle, still, walking), which in turn influences the order of the message topics presented to the user in the auditory flow; (2) Entirely auditory, accessible interfaces that leverage the BVI's auditory bandwidth to attend to and rapidly process aural prompts; (3) Screenless input techniques that leverage finger-worn devices to discreetly control, browse and manipulate text, thus bypassing the reliance on a visual display.

7.1 Dataset and their Associated Situations

To generate candidate messages, we introduce a dataset of interpersonal text conversations based on a study conducted with 20 BVIs about their needs and practices while on the go. The dataset divides more than 850 text messages into 17 specific topics and maps them to associated situations learned from the study. In order for the user to personalize a message, we define seven parametric message instances that can replace a word of a message with alternative values using a word prediction algorithm. This process yields more than 250k text instances.

We believe that the resulting dataset and the associated situations can shed light on the experience of BVIs using current technologies to send text messages rapidly and to circumvent the difficulty of using multiple gestures while texting on the go. This will open opportunities for new assistive technologies that can minimize the interaction overhead and maximize the communication throughput. While the dataset can be useful to future research in accessibility, we argue that our dataset can also benefit sighted users in other scenarios, such as when the user is driving and cannot hold the phone in their hand. In the future, we plan to perform a study with both blinds and sighted users by simulating a real-world scenario to better assess and compare the usability and user experience with TextFlow across the two groups of users.

7.2 Interaction with Finger-worn Device

In order for the user to interact with the wearable device (TapStrap) using tapping gestures, we define two different interaction modalities: topic-by-topic browsing (TBT) and self-disclosing flow (SDF). TBT interaction modality allows users to browse the system's suggestions with explicit forward and backward commands, whereas the SDF interaction modality reads the system's suggestions sequentially and allows users to make a selection upon hearing the intended option in the auditory list. The results of our evaluative study show that when the system is in the TBT condition users are able to interact fluidly with the system in most instances with a smaller error rate, whereas in the SDF condition they encounter more difficulties and have a higher error rate. This confirms prior work that suggests BVI users are used to have fine-grained control over a list of items on a webpage or content headers [25, 26, 43]. However, in both interaction modalities as users interact more with the system the learnability improves, suggesting that adding a training session in which users can become familiar with the list of topics and messages can further reduce the interaction error.

Based on the analysis of our experiential themes, while TBT interaction modality is preferred over SDF, participants still mentioned situations in which the SDF interaction modality could be useful (e.g., when users are multi-tasking) as well as design suggestions, such as adding a sound (e.g., beep) between topic/message options. In addition, for both interaction modalities participants expressed the desire to access the most frequently used topics/messages at the beginning of the aural stream. We believe this will minimize the amount of time users need to send a text message. In the future, we plan to change the order of the topics/messages based on the associated situations and their frequency, as well as user history. In addition, other analyses demonstrate that the wearable device can support the different combination of hand placement (e.g., leg) and finger position (e.g., flat) without affecting the user performance.

7.3 Design Consideration

The overall experiential response of participants toward TextFlow is positive. Users emphasized that the system bypasses the limitations of existing technologies in terms of reliability, time efficiency, and privacy, in particular when users are mobile and need to send a quick text message while focusing on their surroundings. However, users pointed out other limitations that provide direction for future work. Participants mentioned the need to maintain awareness about their surroundings while texting on the go. While existing assistive technologies for navigation (e.g., Be My Eyes or Seeing AI) allow users to receive information about their vicinity, users still have to switch between multiple applications when sending a text. This multi-tasking process can be both time-consuming and distracting. We are working on incorporating warning cues from the surrounding environment into the auditory TextFlow to increase safety in nomadic contexts. Another limitation is the extent of personalization. Participants expressed that going through all the message options could be time-consuming and increase the cognitive load. Future work will suggest personalized topics and messages, such that their order is changed based on the user history and their situation. In addition, we will provide the option to add custom messages based on users' preferences. A major challenge is to strike a balance between message options suggested by the system and the complexity of text editing when manipulating strings in screenless, aural environments.

ACKNOWLEDGMENTS

The authors would like to thank anonymous reviewers for their insightful comments and suggestions on earlier versions of this manuscript. This research is based on work supported by the National Science Foundation under Grant IIS #1909845. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the NSF. We thank all study participants and the staff at BOSMA Enterprises for assisting with the recruiting activities.

REFERENCES

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. 2018. " Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 249–258.

[2] Ali Abdolrahmani, Ravi Kuber, and Amy Hurst. 2016. An empirical investigation of the situationally-induced impairments experienced by blind mobile device users. In *Proceedings of the 13th Web for All Conference*. 1–8.

[3] Gregory D Abowd, Agathe Battestini, and Thomas O'Connell. 2002. The Location Service: A framework for handling multiple location sensing technologies. *College of Computing and GVU Center, Georgia Institute for Technology, Atlanta,. Georgia, USA* (2002).

[4] Gediminas Adomavicius and Alexander Tuzhilin. 2011. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 217–253.

[5] Tousif Ahmed, Roberto Hoyle, Kay Connelly, David Crandall, and Apu Kapadia. 2015. Privacy concerns and behaviors of people with visual impairments. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3523–3532.

[6] Shiri Azenkot and Nicole B Lee. 2013. 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*. 1–8.

[7] Linas Baltrunas, Bernd Ludwig, Stefan Peer, and Francesco Ricci. 2012. Context relevance assessment and exploitation in mobile recommender systems. *Personal and Ubiquitous Computing* 16, 5 (2012), 507–526.

[8] Nikola Banovic, Koji Yatani, and Khai N Truong. 2013. Escape-keyboard: A sight-free one-handed text entry method for mobile touch-screen devices. *International Journal of Mobile Human Computer Interaction (IJMHCI)* 5, 3 (2013), 42–61.

[9] Victoria Bellotti, Bo Begole, Ed H Chi, Nicolas Ducheneaut, Ji Fang, Ellen Isaacs, Tracy King, Mark W Newman, Kurt Partridge, Bob Price, et al. 2008. Activity-based serendipitous recommendations with the Magitti mobile leisure guide. In *Proceedings of the sigchi conference on human factors in computing systems*. 1157–1166.

[10] Jan Blom, Anu Kankainen, Tomi Kankainen, and Sauli Tiitta. 2003. Location-aware multiuser messaging: exploring the evolution of mobile text-based communication services. *Helsinki Institute for Information Technology Technical Report 2003* 2 (2003).

[11] Matthew N Bonner, Jeremy T Brudvik, Gregory D Abowd, and W Keith Edwards. 2010. No-look notes: accessible eyes-free multi-touch text entry. In *International Conference on Pervasive Computing*. Springer, 409–426.

[12] Ourdia Bouidghaghen, Lynda Tamine-Lechani, and Mohand Boughanem. 2009. Dynamically personalizing search results for mobile users. In *International Conference on Flexible Query Answering Systems*. Springer, 99–110.

[13] Djallel Bouneffouf, Amel Bouzeghoub, and Alda Lopes Gançarski. 2012. A contextual-bandit algorithm for mobile context-aware recommender system. In *International Conference on Neural Information Processing*. Springer, 324–331.

[14] Djallel Bouneffouf, Amel Bouzeghoub, and Alda Lopes Gançarski. 2012. Following the user's interests in mobile context-aware recommender systems: The hybrid-E-greedy algorithm. In *2012 26th International Conference on Advanced Information Networking and Applications Workshops*. IEEE, 657–662.

[15] Stephen Brewster, Joanna Lumsden, Marek Bell, Malcolm Hall, and Stuart Tasker. 2003. Multimodal'eyes-free'interaction techniques for wearable devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 473–480.

[16] Maria Claudia Buzzi, Marina Buzzi, Barbara Leporini, and Amaury Trujillo. 2014. Designing a text entry multimodal keypad for blind users of touchscreen mobile phones. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. 131–136.

[17] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. " What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[18] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[19] Android Documentation. 2020. Activity Recognition API. https://developers.google.com/location-context/activity-recognition

[20] Android Documentation. 2020. android.speech.tts Reference page. https://www.tapwithus.com

[21] Android Documentation. 2020. Google Calendar API. https://developers.google.com/calendar

[22] Android Documentation. 2020. Location and Context APIs. https://developers.google.com/location-context

[23] W Keith Edwards and Rebecca E Grinter. 2001. At home with ubiquitous computing: Seven challenges. In *International conference on ubiquitous computing*. Springer, 256–272.

[24] Andrew Fowler, Kurt Partridge, Ciprian Chelba, Xiaojun Bi, Tom Ouyang, and Shumin Zhai. 2015. Effects of language modeling and its personalization on touchscreen typing performance. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 649–658.

[25] Mikaylah Gross and Davide Bolchini. 2018. Beyond screen and voice: augmenting aural navigation with screenless access. *ACM SIGACCESS Accessibility and Computing* 121 (2018), 1–1.

[26] Mikaylah Gross, Joe Dara, Christopher Meyer, and Davide Bolchini. 2018. Exploring Aural Navigation by Screenless Access. In *Proceedings of the 15th International Web for All Conference*. 1–10.

[27] Tiago Guerreiro, Paulo Lagoá, Pedro Santana, Daniel Gonçalves, and Joaquim Jorge. 2008. NavTap and BrailleTap: Non-visual texting interfaces. In *Proc. Rehabilitation Engineering and Assistive Technology Society of North America Conference (Resna)*.

[28] Tiago Guerreiro, Hugo Nicolau, Joaquim Jorge, and Daniel Gonçalves. 2009. NavTap: a long term study with excluded blind users. In *Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility*. 99–106.

[29] Tapstrap Inc. 2016. Tap Strap 2. https://www.tapwithus.com

[30] Robin Jia and Percy Liang. 2017. Adversarial examples for evaluating reading comprehension systems. *arXiv preprint arXiv:1707.07328* (2017).

[31] Simon Jones and Eamonn O'Neill. 2009. Context-aware messaging: how personal, spatial and temporal constraints affect text-based communication. In *Proceedings of the 8th International Conference on Mobile and Ubiquitous Multimedia*. 1–10.

[32] Shaun K Kane, Jeffrey P Bigham, and Jacob O Wobbrock. 2008. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility*. 73–80.

[33] Anjuli Kannan, Karol Kurach, Sujith Ravi, Tobias Kaufmann, Andrew Tomkins, Balint Miklos, Greg Corrado, Laszlo Lukacs, Marina Ganea, Peter Young, et al. 2016. Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 955–964.

[34] Hsin-Liu Kao, Artem Dementyev, Joseph A Paradiso, and Chris Schmandt. 2015. NailO: fingernails as an input surface. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 3015–3018.

[35] Pegah Karimi, Emanuele Plebani, and Davide Bolchini. 2021. Textflow: Screenless Access to Non-Visual Smart Messaging. In *26th International Conference on Intelligent User Interfaces*. 186–196.

[36] Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. 2019. Natural questions: a benchmark for question answering research. *Transactions of the Association for Computational Linguistics* 7 (2019), 453–466.

[37] Barbara Leporini, Maria Claudia Buzzi, and Marina Buzzi. 2012. Interacting with mobile devices via VoiceOver: usability and accessibility issues. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*. 339–348.

[38] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461* (2019).

[39] Wei Li, Xuerui Wang, Ruofei Zhang, Ying Cui, Jianchang Mao, and Rong Jin. 2010. Exploitation and exploration in a performance based contextual advertising system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. 27–36.

[40] Xin Liu and Karl Aberer. 2013. SoCo: a social network aided context-aware recommender system. In *Proceedings of the 22nd international conference on World Wide Web*. 781–802.

[41] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).

[42] Joanna Lumsden and Andrew Gammell. 2004. Mobile note taking: Investigating the efficacy of mobile text entry. In *International Conference on Mobile Human-Computer Interaction*. Springer, 156–167.

[43] Darren Lunn, Simon Harper, and Sean Bechhofer. 2011. Identifying behavioral strategies of visually impaired users to improve access to web content. *ACM Transactions on Accessible Computing (TACCESS)* 3, 4 (2011), 1–35.

[44] Micol Marchetti-Bowick and Nathanael Chambers. 2012. *Learning for microblogs with distant supervision: Political forecasting with twitter*. Technical Report. MICROSOFT CORP SAN FRANCISCO CA.

[45] Estefanía Martín, Pablo A Haya, and Rosa M Carro. 2013. Adaptation technologies to support daily living for all. In *User modeling and adaptation for daily routines*. Springer, 1–21.

[46] Reeti Mathur, Aishwarya Sheth, Parimal Vyas, and Davide Bolchini. 2019. Typing Slowly but Screen-Free: Exploring Navigation over Entirely Auditory Keyboards. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 427–439.

[47] David E Meyer and David E Kieras. 1997. A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms. *Psychological review* 104, 1 (1997), 3.

[48] Sachi Mizobuchi, Mark Chignell, and David Newton. 2005. Mobile text entry: relationship between walking speed and text input task difficulty. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*. 122–128.

[49] Myo. 2018. Welcome to Myo™ Support. https://support.getmyo.com/hc/en-us

[50] Nikita Nangia, Adina Williams, Angeliki Lazaridou, and Samuel R Bowman. 2017. The repeval 2017 shared task: Multi-genre natural language inference with sentence representations. *arXiv preprint arXiv:1707.08172* (2017).

[51] Andronikos Nedos, Alexander O'Connor, Graham Abell, Siobhán Clarke, and Vinny Cahill. 2004. LATTE: Location and Time Triggered Email.. In *International Conference on Wireless Networks*. Trinity College Dublin, Dublin, Ireland, 813–819.

[52] Hugo Nicolau, Tiago Guerreiro, Joaquim Jorge, and Daniel Gonçalves. 2010. Proficient blind users and mobile text-entry. In *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*. 19–22.

[53] Mark Perry and C Shangar. 2005. Marked Spaces: interaction design and sensor integration for ubiquitous location-based messaging and communication. In *3rd UK-UbiNet Workshop: Designing, evaluating and using ubiquitous computing systems*.

[54] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. Improving language understanding by generative pre-training. (2018).

[55] André Rodrigues, Kyle Montague, Hugo Nicolau, and Tiago Guerreiro. 2015. Getting smartphones to talkback: Understanding the smartphone adoption process of blind users. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. ACM, New York, NY, USA, 23–32.

[56] Kristen Shinohara and Jacob O Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 705–714.

[57] Radu-Daniel Vatavu. 2017. Visual impairments and mobile touchscreen interaction: state-of-the-art, causes of visual impairment, and design guidelines. *International Journal of Human–Computer Interaction* 33, 6 (2017), 486–509.

[58] Michael Völske, Martin Potthast, Shahbaz Syed, and Benno Stein. 2017. Tl; dr: Mining reddit to learn automatic summarization. In *Proceedings of the Workshop on New Frontiers in Summarization*. 59–63.

[59] Hao Wang, Zhengdong Lu, Hang Li, and Enhong Chen. 2013. A Dataset for Research on Short-Text Conversations Proceedings of the Conference on Empirical Methods in Natural Language Processing. ACL.

[60] Youdao. 2020. Translator. http://fanyi.youdao.com