Voicemoji: Emoji Entry Using Voice for Visually Impaired People

Mingrui "Ray" Zhang
The Information School,
University of Washington
Seattle, WA
mingrui@uw.edu

Ruolin Wang
UCLA HCI Research,
Los Angeles, CA
violynne@ucla.edu

Xuhai Xu
The Information School,
University of Washington
Seattle, WA
xuhaixu@uw.edu

Qisheng Li
Paul G. Allen School of Computer
Science & Engineering,
University of Washington
Seattle, WA
liqs@cs.washington.edu

Ather Sharif
Paul G. Allen School of Computer
Science & Engineering,
University of Washington
Seattle, WA
ather@cs.washington.edu

Jacob O. Wobbrock
The Information School,
University of Washington
Seattle, WA
wobbrock@uw.edu

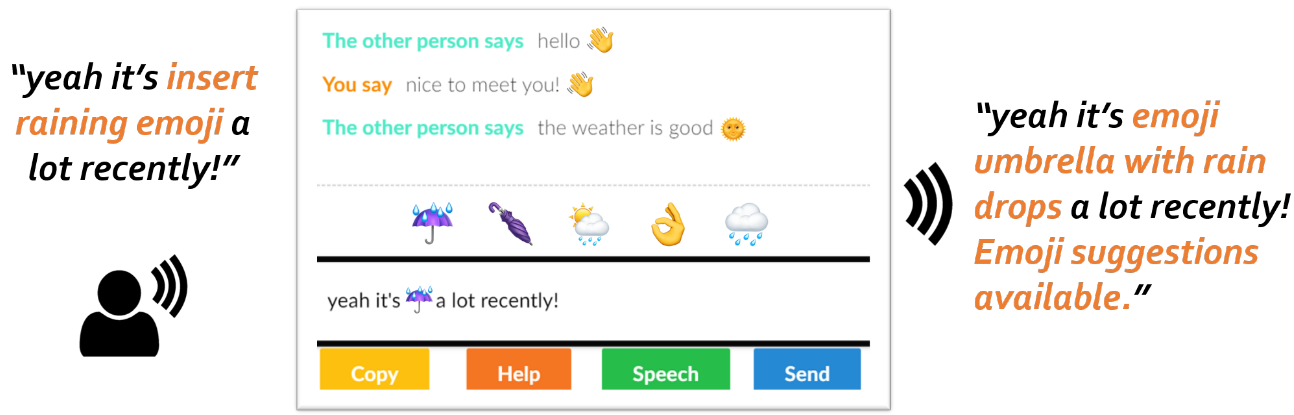

Figure 1: The flow of using Voicemoji. Voicemoji is a web application that allows the user to speak text and emojis. It also provides context-sensitive emoji suggestions based on the spoken content.

ABSTRACT

Keyboard-based emoji entry can be challenging for people with visual impairments: users have to sequentially navigate emoji lists using screen readers to find their desired emojis, which is a slow and tedious process. In this work, we explore the design and benefits of emoji entry with speech input, a popular text entry method among people with visual impairments. After conducting interviews to understand blind or low vision (BLV) users' current emoji input experiences, we developed *Voicemoji*, which (1) outputs relevant emojis in response to voice commands, and (2) provides context-sensitive emoji suggestions through speech output. We also conducted a multi-stage evaluation study with six BLV participants from the United States and six BLV participants from China, finding that Voicemoji significantly reduced entry time by 91.2% and was preferred by all participants over the Apple iOS keyboard. Based on our findings, we present Voicemoji as a feasible solution for voice-based emoji entry.

CCS CONCEPTS

• **Human-centered computing** → **Natural language interfaces**; **Accessibility systems and tools**; **Text input**.

KEYWORDS

Voice-based emoji input, speech user interfaces, accessibility, blind or low vision users, visual impairments.

ACM Reference Format:

Mingrui "Ray" Zhang, Ruolin Wang, Xuhai Xu, Qisheng Li, Ather Sharif, and Jacob O. Wobbrock. 2021. Voicemoji: Emoji Entry Using Voice for Visually Impaired People. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3411764.3445338

1 INTRODUCTION

Emojis have become an essential element of online communication, with over 3,000 emojis available in the Unicode standard [18].

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05...$15.00
https://doi.org/10.1145/3411764.3445338

Facial expressions, emotions, activities, and objects are succinctly represented using emojis. Emojis are widely used in everyday social interactions including text messaging, posting on social media, contacting customer service, and appealing to online audiences through advertisements, making emojis undoubtedly a popular and important way of communicating in today's digital age [46, 63].

The prevalence of emojis in online communications means that blind or low vision (BLV) users encounter emojis often. According to a recent study by Tigwell *et al.* [55], 93.1% of BLV users encounter emojis each month, and 82.7% of them utilize emojis at least once a month. However, due to emojis' similarity to images and the lack of accessibility support for screen readers [55], current emoji entry methods, including *emoji keyboards*, *emoji shortcuts*, and *built-in emoji search*, are cognitively demanding and unreasonably time-consuming for BLV users. We compare current emoji entry methods and summarize their shortcomings, including *emoji keyboards*, *emoji shortcuts*, and *built-in emoji search*, in Table 1 and Figure 2.

Such limitations hinder BLV users from using emojis easily, causing social exclusion for BLV users, and reducing their communication efficacy [57]. Through our interviews with BLV users (N=12), we report that there are four major challenges of current emoji entry methods: (1) the entry process is time-consuming; (2) the results provided by the methods are not consistent with users' expectations; (3) there is a lack of support for discovering new emojis; and (4) there is a lack of support for finding the right emojis. In summary, the current state of searching for and inputting emojis for BLV users is inaccessible, tedious, and exclusionary.

Prior work has reported that BLV users employ voice commands more frequently, and are more satisfied with speech recognition, than sighted people [3]. Gboard has support for dictating emojis, where commands like "fire emoji" would input the 🔥 emoji [64]. Apple voice control allows emojis with multi-word descriptions to be inputted by using similar commands [31]. However, both methods only work when the user knows the exact name of the emoji.

We present *Voicemoji* (Figure 1), a voice emoji entry system that supports: (1) voice-based semantic-level emoji search, (2) emoji entry with keywords, (3) context-sensitive emoji suggestions, and (4) manipulation of emojis with voice commands, such as changing the emoji color or skin tone. Specifically, Voicemoji detects a set of keywords to trigger the emoji input function, and utilizes the results from the Google search engine to find the most relevant emojis. Powered by deep learning, it also suggests emojis based on the spoken content. With Voicemoji, the user can use ambiguous descriptions, such as, "ocean animal emoji," to get a group of emojis including squid 🦑, octopus 🐙, and tropical fish 🐠. Following a similar approach, exploration and learning of new emojis is also possible, which is exceptionally difficult with current emoji input methods.

Additionally, Voicemoji, at present, supports a rich emoji set accessible through two of the three most spoken languages in the world,[1] Chinese and English. This feature enhances the generalizability of our solution in two respects: (1) language independence (*i.e.*, the method can apply to multiple languages); (2) emoji independence (*i.e.*, the method can output all emojis in the current emoji set). We also open-sourced our code to support the research community and provide a platform for contributions from like-minded researchers and developers.[2]

We conducted a multi-stage study to evaluate Voicemoji with six BLV participants from the United States and six BLV participants from China. After learning the usage of Voicemoji in an initial training session, participants were encouraged to use the Voicemoji system in their daily chat conversations for three days. Then, they participated in a lab study to compare the performance of Voicemoji with their current keyboard-based emoji entry system.

Our results show that participants entered emojis significantly faster with Voicemoji than with the Apple iOS keyboard, and the suggestion function of Voicemoji was perceived as relevant and helpful. Qualitative analysis shows evidence that Voicemoji not only improved the emoji entry experience, but also enriched participants' overall online communication experience.

We make three primary contributions in this work:

(1) Through semi-structured interviews, we report on the current emoji input experiences and challenges faced by BLV users;
(2) We developed *Voicemoji*, a speech-based emoji entry system that enables BLV users to input emojis. We contribute its interaction design, including its commands, functionality, and feedback, which support a multilingual system. Additionally, we provide the source code of our implementation;
(3) Through a multi-stage user study, we evaluated the usability of Voicemoji and compared it to current emoji entry methods. Our results show that Voicemoji significantly reduces input time for emoji entry by 91.2% and is highly preferred by users.

2 RELATED WORK

Prior research related to the current work exists in a variety of areas, which we review in this section. These areas are the role of emojis in online communication, making emojis accessible to blind or low vision (BLV) users, and the use of speech-based interfaces among BLV users. We cover each of these in turn.

2.1 The Role of Emojis in Online Communication

Emojis are a set of pictorial Unicode characters with visual representations of expressions, activities, objects, and symbols. Since they were inducted into the Unicode Standard in 2009 [16], the usage of emojis has increased dramatically. A 2015 report by Swiftkey [53] revealed that users inputted over a billion emojis over four months; a similar report in 2017 from Emojipedia [8] showed that five billion emojis were sent *daily* on Facebook messenger. Because of their pictorial appearance, emojis "convey information across language, culture, lifestyle and diversity" [1]. In fact, people sometimes even use pure emoji combinations unaccompanied by text to covey their expressions (*e.g.,* 📚✈️ = *book a flight*) [12, 32].

People use emojis for different purposes. Emojis can be used to provide additional emotional context or situational information [15], change the tone of a message, engage the recipient, or maintain

[1] https://www.babbel.com/en/magazine/the-10-most-spoken-languages-in-the-world

[2] https://github.com/DrustZ/VoiceEmoji

Table 1: Summary of different emoji entry methods. Voicemoji aims to address several problems of current methods by providing features including voice input, fuzzy semantic-level search and emoji suggestions

Method	Modality	Emoji search?	Emoji suggestions?
Emoji keyboard	Touch	No	No
Emoji shortcut	Touch	No	No
Built-in emoji search	Touch	Keyword-level	No
Voicemoji	Voice	Keyword- and semantic-level search	Semantic-level

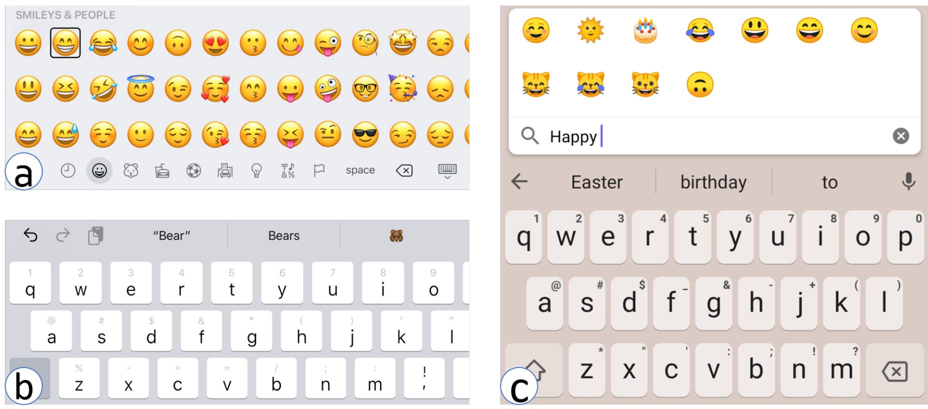

Figure 2: Current emoji input methods. (a) The Apple iOS emoji keyboard. Blind or low vision users can use a finger to move over the keyboard to navigate through the options one by one, and the screen reader will read out the name of each emoji. The process is tedious and slow. (b) The emoji shortcut. When the user types certain keywords, such as "bear," the corresponding emojis will appear in the suggestion list. Not all emojis have shortcuts, and the user has to memorize the shortcuts to find them. (c) Built-in emoji search. Some keyboards offer built-in search functions, where the user can input text to search for emojis. The search is only based on manually curated keywords. In our interviews, we found that our participants mainly used the emoji keyboard (a) to input emojis, despite its evident drawbacks.

a relationship [15, 34]. People also use emojis in highly personalized and contextualized ways to create "shared and secret uniqueness" [34, 48, 54, 65]. An example provided by Wiseman and Gould [65] showed that a romantic couple used the pizza emoji 🍕 to mean "I love you" because of their shared love for pizza. In general, the usage of emojis improves the expressiveness of online communication [29, 33, 59].

It is not surprising that people who are blind or have low vision (BLV) also use emojis in their written communications. According to Tigwell *et al.* [55], over 93.1% BLV users engage with emojis at least once a month. These people's purposes when using emojis are the same as those of sighted people, including enhancing message content, adding humor, and altering tone. Unfortunately, the accessibility and usability of emoji interfaces for BLV users is lacking, although there have been some efforts to improve upon this situation. We now review these efforts.

2.2 Emoji Accessibility for Blind and Low Vision Users

Visual content such as images, videos, stickers, badges, memes, and emojis can enhance online communication and social interaction. Unfortunately, much of this visual content remains inaccessible. Prior work mainly has focused on improving the accessibility of pictures [38, 42] and memes [20, 21, 47] posted on social media. The work has utilized human-in-the-loop plus automatic methods such as optical character recognition and scene description. In the last decade, emojis have become a staple in online communication [35, 70]. Although a large body of work on emojis has explored the inclusiveness of emojis along the lines of gender [5, 7, 10], age [27], and race [4, 9], making emojis accessible for different user groups is still an open research topic.

Owing to their inaccessible pictorial nature, emojis can be easily misunderstood by BLV users. For example, the same emoji can have different definitions on different platforms, and can also be read differently by different screen readers. This inconsistency can cause frustration and misunderstanding [55]. Furthermore, many emojis have similar descriptions, such as 😄 (*Grinning Face with Smiling Eyes*) and 😊 (*Smiling Face with Smiling Eyes*), which are hard for a person to distinguish without visual portrayals. As a consequence, research shows BLV users can lack confidence when selecting emojis [55]. To help remedy this problem, Kim *et al.* [36] combined machine learning and k-means clustering to analyze the conversation and recommend emojis that represent various contexts, which can ease the challenge of selecting appropriate emojis for BLV users.

Outside some prior academic research, there has been little recognition in industry that the inaccessibility of emojis is a problem. As noted above, the predominant way to select and input an emoji is to visually search over an emoji keyboard, which offers emojis in a multi-page menu, grouped by theme. This method is imprecise and slow, even for a sighted user; it is even more unusable for blind or low vision (BLV) users. A video on YouTube [3] demonstrates the procedure vividly: to navigate among emoji options, a BLV user has to make a swipe or drag the finger to move the focus onto the next emoji, and continue to step until the expected emoji is spoken (Figure 2a).

Recently, productized mobile keyboards have added emoji suggestions in the suggestion bar when certain keywords are typed. For example, typing "bear" results in a bear emoji 🐻 appearing (Figure 2b). Although such shortcut suggestions reduce user workload, users still have to memorize corresponding keywords, and emojis with complex descriptions, such as 😸 (*Grinning Cat Face with Smiling Eyes*), cannot be inserted in this way, as they do not have a short keyword. Lastly, emoji keyboards such as *Gboard* have built-in emoji search functions (Figure 2c). However, the emoji search function is based on keywords that are assigned manually; hence, this function does not provide much flexibility. For example, users can enter "fruit" to find fruit emojis, but "fruits" produces zero results. Another issue of keyword-based search is the poor transferability between languages: searching emojis on a Chinese keyboard often leads to fewer results than performing the same search on an English keyboard because the names of emojis are defined in English, and there are no official emoji names in Chinese.

That said, both academic researchers and industry developers have put effort into improving the accessibility of emoji systems. Web developers utilized the Aria-label with emoji text to standardize the description for screen readers [51]. Researchers also designed emojis [11] that can be combined with Braille text. And prior work has addressed the problem of how to make emoji *output* more accessible [14, 61]; however, there has been precious little work on making emoji *input* more accessible, which is the focus of this work.

2.3 Speech Interfaces for Blind or Low Vision Users

Speech-based interfaces are commonly used to support accessibility (*e.g.*, [17]), and even some non-speech voice-based accessible applications have been developed (*e.g.*, [23]). Research shows that screen reader users are generally more satisfied with speech input than are non-screen reader users [3]. Speech input is not only faster than typing [49], the performance of state-of-the-art recognition algorithms has also reached the word-level accuracy over 95% [45]. Voice input has also been applied to other tasks such as "free-hand" drawing [24, 30] and image manipulation [26].

As a consequence, when considering solutions for BLV users to enter emojis, speech-based interfaces are a natural consideration. We are not the first to consider a "speak emoji" design; in fact, *Gboard* has enabled a function to speak single-word emojis in its speech-to-text engine [5]. For example, when the user speaks "fire emoji," Gboard will transcribe the speech into 🔥. However, it only allows single-word emoji entry, and users have to know the exact keyword for the desired emoji. The Apple iOS voice control[4] also allows a user to input emojis using voice; however, the function was designed specifically for people with motor impairments, and the user has to memorize the exact keyword of the emoji they desire. Moreover, all of the current voice emoji interfaces do not support exploration: they only show one emoji according to a single-word spoken command.

Unlike current solutions, Voicemoji presents a comprehensive set of approaches to speak emojis: the user can either input an emoji with a key phrase, or find relevant emojis by natural language commands. Voicemoji also provides emoji suggestion features inspired by Emojilization [28], which provide emoji suggestions for speech input based on the semantic meaning of the spoken text and the tone of the speech. This helps the user discover unfamiliar emojis and find proper ones to avoid misuse [55].

3 UNDERSTANDING CURRENT EMOJI USAGE BY BLIND USERS

To design an emoji entry method for blind or low vision (BLV) users, we first need to understand the problems they face when they utilize current emoji systems. To gain this understanding, we conducted multiple semi-structured interviews[5] with our target users from both the United States and China. Specifically, we wanted to answer three questions: (1) How do BLV users currently input emojis? (2) How do BLV users discover and conceive of new emojis? (3) What are the main challenges of using the current emoji entry methods for BLV users?

We recruited 12 participants, 6 from the United States (5 men, 1 woman) aged 18 to 68 (M=35.5, SD=17.1), and 6 from China (4 men, 2 women) aged 25 to 27 (M=26.0, SD=0.9). We contacted our participants by sending emails to BLV community centers. Our participants' demographic information is shown in Table 2. Seven participants identified as blind and five participants identified as low vision. All participants owned mobile phones and used them daily with screen readers. Due to the inconsistency of emoji descriptions on different platforms, we only recruited Apple iOS users, as the iOS system has more detailed descriptions for each emoji than Android in both English and Chinese. The interviews lasted 45-60 minutes and were audio-recorded for analysis. Our interview protocol was guided by our research questions. Participants were compensated with $15 USD or 100 CNY for their time.

For analysis, two authors independently coded all of the interview transcripts while discussing and modifying the codebook to reconcile ambiguities on an ongoing basis. The research team discussed any discrepancies until reaching consensus. We did not, however, calculate inter-rater reliability, as the primary goal of the coding process was not to achieve complete agreement, but to eventually yield overarching themes [40]. After coding all interviews, all authors conducted multiple sessions of thematic analysis of the interviews, using affinity diagramming [50] as a modified

[3] https://youtu.be/PpqLnO-1Kxw

[4] https://support.apple.com/en-us/HT210417

[5] Due to the COVID-19 pandemic, we conducted all interviews online.

Table 2: Demographic information of participants

ID	Sex	Age	Nation	Visual Impairment	Frequency of Using Emojis
P1	F	25	CN	Blind	Every day
P2	M	25	CN	Blind	Every day
P3	M	26	CN	Low Vision (Retinitis pigmentosa)	Every day
P4	F	26	CN	Low Vision (Retinitis pigmentosa)	Every day
P5	M	27	CN	Low Vision (Retinitis pigmentosa)	Rarely use them
P6	M	27	CN	Low Vision (Peripheral vision loss)	Every day
P7	M	30	US	Blind	Once a week
P8	F	68	US	Low Vision (Central vision loss)	Every day
P9	M	18	US	Blind	Once a week
P10	M	35	US	Blind	Every day
P11	M	35	US	Blind	Rarely use them
P12	M	27	US	Blind	Every week

version of grounded theory [13] to uncover themes of various levels. We present our results in the following subsections. Some participant quotes have been edited slightly and shortened to improve readability without changing their substance.

3.1 Current Emoji Entry Practice

We briefly report participants' current emoji entry practices in this section.

Frequency and Motivation. Eight participants reported that they used emojis every day, two used emojis once a week, and two rarely used them (*i.e.*, less than once a week). This result aligns with previous work [55] indicating the popular usage of emojis despite their pictorial and visual nature. However, we found that most participants only used a limited number of emojis frequently (about 10), most of which were emotion-related ones such as smiling faces. When asked about motivations, all people mentioned *enriching the expressiveness* of their communications; one participant (P10) also mentioned using emojis as a quick response. Interestingly, two participants (P8, P10) also mentioned the sense of *belonging and connecting to their peers* when using emojis, which emphasized the social aspects of emoji usage.

Input Methods. For daily communication, six participants used speech input as their main text input method, three used an on-screen keyboard, and three used a braille keyboard. For those who did not use speech input as their main method, they all used speech input for certain situations, such as "quick stuff" (P12) or "when I'm lazy" (P7). All participants reported using the emoji keyboard as the main input method for emojis. For those who used emojis frequently, they would memorize the position of certain emojis. Five participants also utilized the *frequently used page* of the emoji keyboard to speed up their input process. Participants also mentioned using emoji shortcuts, but only P2 used them as the main way to input emojis, memorizing the keywords that brought up certain emoji suggestions. However, other participants felt that this approach was too unpredictable, and they often did not know what keyword could provide a desired emoji suggestion.

Learning New Emojis. To discover new emojis, seven participants mentioned that they got to know new emojis while they were swiping to input a known emoji on the emoji keyboard. Five participants occasionally scrolled through the whole emoji keyboard. Six participants mentioned discovering new emojis from the messages sent to them. Two people also read release notes, such as Unicode specifications, to learn new emojis. Everyone learned new emojis by their emoji descriptions; however, P10 mentioned that a lot of these descriptions were "confusing and not detailed enough." Participant 8 mentioned that she would connect her phone with a television magnifier to see new emojis, and P12 would search on *Emojipedia.org* to learn new emojis. Five participants also mentioned that they would consult with their sighted friends about how to use certain emojis to feel confident using them. Although all of these learning methods exhibit the tenacity and cleverness of our participants, they amount to labor-intensive workarounds that should be avoidable with better designs.

3.2 Challenges of Current Emoji Entry Methods

Through our interviews, we identified several problems with current emoji entry methods for blind or low vision (BLV) users. Many of these problems reduce usability for non-BLV users as well, and unsurprisingly, improving emoji systems for BLV users is likely to improve emoji systems for all users.

C1. Time Consuming. All participants complained about the inconvenience of using the emoji keyboard. There are thousands of options in the Apple iOS emoji keyboard, and these options are even grouped by categories and similarity; but it is still time-consuming to listen to the description of each emoji one-by-one. Participant 7 said, "I actually have to read every single one on the page to know what's on that page. So it just it's time consuming." Participant 2 mentioned that when the procedure was too long she would just give up. The emoji shortcut method was faster compared to the emoji keyboard, but "typing and correcting the text still took time" (P2, P7). Participant 1 and P10 also mentioned that when using the shortcut method, they have to think about the keyword and might try different words to trigger an emoji suggestion.

C2. Inconsistent with Users' Expectations. The main challenge of the emoji shortcut method was the lack of consistency with users' expectations. There was no guarantee that every keyword would result in an emoji suggestion, and the user had to guess the right keywords. Participant 7 said, "When it works, it works really well. When it doesn't work, I have to guess several times and if all the keywords fail, I'm pretty confused." Participant 4 also expressed her confusion: "Sometimes I type exactly the description of the emoji but it does not show the suggestion. Then I feel it is stupid and do not know what to type." The timing of the emoji suggestions was also inconsistent. Some emoji suggestions appeared as auto-correction candidates, while others appeared as auto-prediction candidates. Participant 3 provided an example: typing "happy birthday" would lead to a partying face emoji suggestion 🥳 in the list, but after the space bar was pressed, the emoji changed to a balloon emoji 🎈. There was also inconsistency in the emoji keyboard method. Many participants mentioned that the categories were ambiguous; for example, the category *Smileys & People* contained cat face emojis.

C3. Lack of Support for Discovering Emojis. There was also no convenient way for BLV users to discover new emojis. Most participants mentioned that they only used a limited number of common emojis. While the emoji list contains many options, only five participants mentioned that they would occasionally navigate among the keyboard to explore new emojis.

C4. Lack of Support for Finding the Right Emoji. Not knowing enough emojis limited the expression participants could convey through emojis. Participant 2 mentioned, "Sometimes I want to add some emojis, but I don't know which to add so I just give up." The keyword suggestion method can mitigate this challenge to some degree, but not always: "If the emoji is suggested by the keyword, I would pick it. However, it might not be the best one in my mind. I pick the suggested ones only because it was too tiring to pick the right one from the emoji keyboard" (P4). Even if users know the emoji exists, they usually ask a sighted friend to explain the context of the emoji. There is no way for them to discover the proper usage context of new emojis with current input methods.

3.3 Features Emerged from the User Interviews

Based on our interview results, we summarize certain key features that *Voicemoji* needs to have to address these challenges.

F1. Support Direct Emoji Entry. To address challenge *C1*, when users have specific emojis in mind, they can directly and easily insert those emojis via speech. Ideally, users can speak both emojis and text in one utterance without explicitly switching modes.

F2. Enable Natural Language Queries. To address challenge *C2*, with Voicemoji, users should be able to ask for emojis in a natural way, rather than having to remember keywords or names. For example, when looking for the emoji *Man with Probing Cane* 👨‍🦯, users can simply say, "a blind person emoji" instead of the whole, exact name.

F3. Offer Various Options Related to the Query. To address challenges *C3* and *C4*, Voicemoji should be able to scope relevant emojis if users are unsure which emojis to use. Scoping the results is recommended by the human-AI interaction guidelines [2], where it is suggested offering the user more options from which to choose, and opportunities to discover new options.

F4. Suggest Emojis Related to the Current Context. To address challenge *C3* and *C4*, when users do not know which emojis to use, Voicemoji needs to provide suggestions based on the current message content.

F5. Provide the Ability for Color or Skin Tone Modification. Besides the major challenges, two participants (P7, P10) explicitly mentioned the difficulty of choosing skin tones for certain emoji. With the Apple iOS emoji keyboard, users have to long-press an emoji to trigger the skin tone selector, and go through extra steps to modify the skin tone of an emoji. For a better user experience, users should be able to specify or modify the color of an emoji with speech directly and easily.

4 THE DESIGN AND IMPLEMENTATION OF VOICEMOJI

In this section, we describe the design and implementation of *Voicemoji* (Figure 1). Voicemoji contains several speech commands to trigger emoji search, emoji insertion, and emoji modification. The list of commands is shown in Table 3.

Voicemoji can be operated using only speech, or in conjunction with VoiceOver or other screen readers. We implemented Voicemoji as a web application for easy cross-device access without the need of app installation. When the user clicks the *speech* button, the screen reader prompts, "Please start speaking," and users can begin their speech input. Users click the button again when they finish speaking. We used the Google Cloud speech-to-text API[6] for speech recognition. To improve the usability of Voicemoji, we added a "copy button" so that users could copy the spoken content and paste it into other messaging apps; we also added a "help" button that announces the basic usage and commands of Voicemoji. To support our remote user study, described in the next section, we added a chat feature in Voicemoji: clicking the "Send" button would send the text to other the users on the website at the same time. The overall process of using Voicemoji is depicted in Figure 3. We next introduce each command for searching and inputting emojis.

4.1 The Emoji Search Command

We designed the command template, "Emoji search: *description* + emoji" to explicitly search for certain emojis. Voicemoji extracts the description between "Emoji search:" and "emoji" as the query, returning related emojis and announcing them with Apple VoiceOver. For example, the user can say, "Emoji search: a blind person emoji," and Voicemoji will return emojis including 👨‍🦯 (*Man with Probing Cane*), 🦯 (*Probing Cane*), and 🦮 (*Guide Dog*). Upon receiving the results, Voicemoji triggers the screen reader to announce the names of the emojis one by one. The emoji results are shown as buttons, and the user can either tap an emoji, or select an emoji by its position by saying, for example, "Insert the second one." The usage flow is shown in Figure 4.

As specified in Section 3.3, the user's search description does not have to be a predefined emoji keyword or name. Voicemoji accepts any form of natural language as the emoji description (feature *F2*, above), such as "tropical fruit" or "cold weather," even though no specific emojis exist with these exact names.

[6]https://cloud.google.com/speech-to-text

Table 3: Emoji input commands of Voicemoji and usage examples

Command	Result	Example command	Example result
"Emoji search: *Description* + emoji"	Return a list of emojis relevant to the description	"Emoji search: A blind person emoji"	👨‍🦯 🦯 🐕‍🦺 👩‍🦯 👨
"Insert + *description* + emoji"	The most relevant emoji is added directly to the transcription	"Happy birthday insert birthday cake emoji"	"Happy birthday 🎂"
"Change the emoji to + *color/skin*"	The last inputted emoji is changed to the corresponding color/skin	(the input text: "that's great! 👍") "Change the emoji to dark skin"	"that's great! 👍🏿"
Emoji suggestion function	Five emoji suggestions relevant to the spoken content when no emoji command is received	"How about dinner tonight?"	🍴 😋 🍝 😏 🍽️

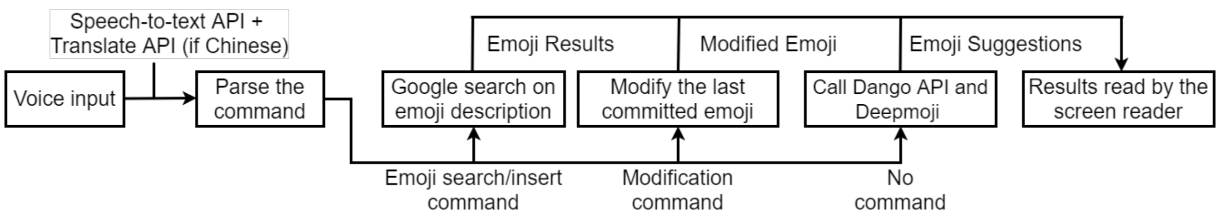

Figure 3: The usage flow of Voicemoji. After the transcript of the voice input is received, the server parses the input to check the type of the command it contains. The parsed input is then processed by different subsystems according to its command type. Finally, emoji results are returned and announced.

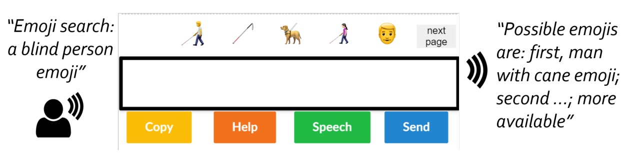

Figure 4: Emoji search command flow. When the user speaks the command, "Emoji search: *description* + emoji," Voicemoji will return related emojis above the text field, and the screen reader announces the name of each returned emoji. If there are more emojis available, the screen reader will read the first five emojis, and then say "more emojis available."

To enable Voicemoji's search functionality, we utilized the Google search API [7] as the search engine to enable flexible search queries with natural language understanding [44]. After extracting the query, Voicemoji searches the query in *Emojipedia* via the API. Emojipedia is an emoji reference website that documents the names of emoji characters in the Unicode Standard. The Google search API finds the most relevant pages in Emojipedia based on the query. Google's search results may contain different types of websites such as blogs, news, and emoji definition pages[8]. Voicemoji then applies regex matching on the resulting pages to extract emoji definition pages, and adds the corresponding emojis into the list of emoji search results (feature *F3*, above). The results are then announced by the device's screen reader. If there are more than five results, a *next page* button will appear to facilitate page navigation.

For the Chinese language, we used similar search commands "给我一个+ *description* + 表情" (表情 stands for "emoji"). After extracting the query, Voicemoji translates it into English using Google's Cloud Translation API[9]. The rest of the procedure is the same as for English search queries.

One essential difference between Voicemoji and searching emojis directly on Google is that Voicemoji is a text-input interaction, and it is targeted at improving the communication efficiency. While the user could get the same result by searching on Google, the portion of the interaction for a visually-impaired user (open a browser, go to Google, search the emoji, go to the website, copy the emoji, switch the application, paste the emoji) is significantly higher than using a built-in Voicemoji-like function from the keyboard.

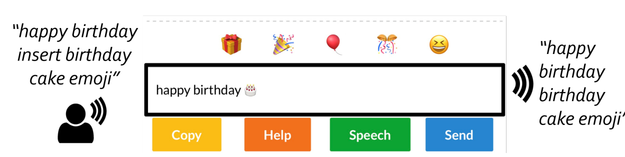

Figure 5: Emoji insertion command flow. When the user speaks the command, "Insert + *description* + emoji," or "*single word description* + emoji," Voicemoji will return the transcribed text with the emoji replacement.

[7]https://developers.google.com/custom-search/v1/overview
[8]For an example of an emoji definition page, see https://emojipedia.org/fire/ for the "fire" emoji 🔥.

[9]https://cloud.google.com/translate

4.2 The Direct Insertion Command

We designed two commands to support direct emoji entry within text (feature *F1*, above). The first one is similar to a feature in Gboard: whenever the user speaks a word followed by the keyword "emoji," Voicemoji will replace the word with a corresponding emoji. For example, saying, "walking in the park with my dog emoji" results in "walking in the park with my 🐕." To avoid replacing words that are actually describing the word "emoji," such as, "I like to use emojis in my daily life," we only trigger the replacement of words that are nouns or gerunds.

The other command is, "Insert + *description* + emoji," and Voicemoji will replace the whole command with the corresponding emoji. This command enables the direct entry of emojis with multi-word descriptions. For example, "Happy birthday insert birthday cake emoji" results in "Happy birthday 🎂." The usage flow is shown in Figure 5. The command for the Chinese language is "插入+ *description* + 表情".

Both the English and Chinese commands replace the description in place, thereby supporting fast emoji entry when the user has a specific emoji in mind. When the processing finishes, the screen reader speaks the transcribed text, including the emoji, to assure the user of the result. The query is processed with the Google search API described above, and the top emoji results are returned.

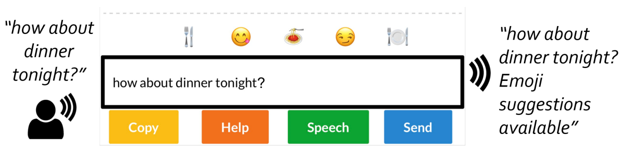

Figure 6: When no emoji command is received, five emoji suggestions are produced by Voicemoji based on the spoken word content. For example, for the phrase, "how about dinner tonight?", Voicemoji produces a fork and knife emoji, smiley face licking its lips emoji, plate of spaghetti emoji, smirking face emoji, and dinner plate with utensils emoji.

4.3 Emoji Suggestions for Spoken Content

When the user does not explicitly ask for emojis during speech, Voicemoji suggests relevant emojis based on the current spoken word content (features *F3, F4*, above). For example, if the user says, "How about dinner tonight?", Voicemoji will return the transcription with suggested emojis including 🍴 (*Fork and Knife*), 🍽️ (*Fork and Knife with Plate*), 🍝 (*Spaghetti*), 😋 (*Smiling Face Licking Lips*) and 😏 (*Smirking Face*). When emoji suggestions are available, the screen reader says, "Emoji suggestions available" after speaking the transcribed result to remind the user. The suggestions are also shown as buttons, and the user could tap to insert them. The usage flow is shown in Figure 6.

This emoji suggestion feature was implemented using two methods: the Dango API [10] and the DeepMoji model [19]. Dango [62] is a mobile application that suggests emojis and stickers based on the message content. DeepMoji is a neural network model trained on 1.2 billion Tweets for emoji prediction. Both methods use neural networks to embed the text into a vector in a semantic space and search for its nearest emoji vectors in the space as the suggestions. For more technical details, the reader is directed to related articles on deep learning and emoji prediction [19, 52]. Voicemoji always returns five suggestions for the current spoken content to improve the emoji variety. After getting the results from the Dango API (usually three or four emojis), Voicemoji then runs the DeepMoji model [11] for further emoji predictions to fill in the remaining slots. We use the Dango API first as it is a commercial product which has a larger training dataset and produces more realistic suggestions than Deepmoji. The suggested emojis reflect both the semantics of a phrase (e.g., suggesting food emojis in the above example) and the affect of a phrase (e.g., suggesting facial expressions). For Chinese input, the content is translated into English and then similarly passed to the Dango API and the Deepmoji model.

[10]https://getdango.com/API/

Figure 7: Color/skin modification usage flow. When the user speaks the command, "Change the emoji to + *skin/color modifier phrase*," Voicemoji will change the last inserted emoji to its corresponding color/skin variation.

4.4 Emoji Modification Commands

Voicemoji also supports modification of already inserted emojis (feature *F5*, above). The user can say the command, "Change the emoji to + *description*" to change an already inputted emoji to another one. For example, if the inputted text is, "Take a walk with my 🐕," speaking "Change the emoji to cat" will modify the dog emoji into a cat emoji 🐈. The command can also modify the color/skin of the emoji if the description contains a color ("yellow," "blue," "green," "brown," "red," etc.) or skin tone ("light," "medium-light," "medium," "medium-dark," "dark," etc.). For example, to change the emoji *thumbs up* 👍 into 👍🏿, the user can say, "Change the emoji to dark skin." The usage flow for this feature is demonstrated in Figure 7.

To implement this color/skin modification feature, we extract the description and decide whether the description contains a color or skin modifier word. If not, Voicemoji just searches the query using the Google search API as usual. If a color or skin modifier word *is* detected, Voicemoji forms a new query by combining the last inserted emoji and the description (for example, 👍 + "dark skin"), and feeds this new query to the Google search API. The first emoji result is then used to replace the inserted emoji in the text.

Voicemoji also supports other modification commands: removing an inserted emoji ("Remove the emoji" or "Delete the emoji"),

[11]https://github.com/bfelbo/DeepMoji

inserting an emoji in the suggestion list ("Insert the first one"), and making the screen reader read the names of emoji options currently in the list ("Read emojis"). All such modification commands make changes to the last inserted emoji. In this way, Voicemoji can be operated entirely via speech if desired.

4.5 Voicemoji Implementation Details

Voicemoji is an web application written in JavaScript. To support its accessibility features, we used aira-labels [41] on the UI elements in order to make them recognizable by a screen reader. The speech output for announcing emoji search results was implemented with the aria-live feature [41]. To facilitate our remote user study, described below, we also used WebSockets to support a real-time messaging function. The Voicemoji backend was implemented with the Tornado library [56]. Finally, the Google Speech, Google Translate, Dango, and DeepMoji services were all incorporated into Voicemoji as described above.

5 VOICEMOJI EVALUATION

We conducted a user study to evaluate Voicemoji and understand its usage. Specifically, we focused on three questions: (1) What is the performance of Voicemoji for emoji entry compared to today's *de facto* entry methods? (2) Is Voicemoji's suggestion feature useful? What is the perceived relevance of the suggested emojis? (3) How does Voicemoji impact the online communication experience of its users? We describe our study to answer these and other questions below.

5.1 Participants

We invited the same 12 blind or low vision (BLV) participants shown in Table 2 from the formative study to this summative evaluation. All of the participants used a mobile phone running the Apple iOS system, and were familiar with the built-in screen reader VoiceOver. The study lasted for about 2 hours, and participants received $30 USD or 200 CNY for their time.

5.2 Apparatus

To optimize for network speed, our testing website was deployed on a university server for our United States participants, and on a commercial Virtual Private Server (VPS) in South Korea for our Chinese participants. We did not deploy the server in China due to the Great Firewall, which blocked the Google services on which Voicemoji relied.

5.3 Procedure

All study sessions were conducted remotely via Zoom or WeChat, and were audio recorded for further analysis. We asked the participants to turn up their VoiceOver volume so that the experimenter could hear the output over their video link. The speech rate of Voice Over was set to 70% for Chinese participants and 65% for American participants, values differing slightly owing to the different output rates of the two languages. The study was conducted on two separate days for each participant. The first phase in the study was a tutorial during which we introduced Voicemoji, and instructed participants how to use different commands for emoji search and selection. Participants tried each available command and demonstrated their ability to use Voicemoji. We then scheduled the summative study three days after the tutorial, and encouraged participants to use Voicemoji for their daily communication needs during the intervening period.

The second phase of the study was a formal summative evaluation, which contained two sessions: the *emoji entry session* and the *emoji suggestion session*, described below.

Emoji Entry Session. We asked the participants to input 27 phrases containing emojis using each emoji entry method (Voicemoji and the current Apple iOS keyboard); the order was counterbalanced. Both methods used speech input for text. (In the iOS keyboard condition, participants were told to input the text with dictation.) When using the iOS keyboard, participants were told they could use their preferred emoji entry method for each phrase. For example, they could use the emoji shortcut method for some phrases and the emoji keyboard for others. The rationale was to regard the iOS keyboard condition as the *de facto* means of inputting emojis, which makes available a combination of different methods, just as in everyday life. Furthermore, participants could always fall back to the emoji keyboard if they could not find the emoji using the shortcuts.

To make the task reflective of real conversational situations, we designed four groups of phrases:

- Five phrases with emojis at the end of the text: "*Are you going to join us for lunch?*🍔 *(hamburger)*"
- Five phrases with emojis replacing a word in the text: "*A* 🎁 *(present) isn't necessary.*"
- Five phrases with emojis at the beginning of the text: "👀 *(eyes) See you soon!*"
- Ten phrases with only emojis: "😂"

We also asked participants to *compose* an original phrase with text plus emojis and a phrase with only emojis, which added up to 27 test phrases in all. Phrase text was randomly selected from the Enron mobile phrase set [58], and the first author added relevant emojis to each phrase. All other authors then reviewed the phrase set and agreed on the relevance of each emoji. For emoji-only phrases, we selected the five most popular emojis at Emoji Stats [12]. The website shows the real-time emoji usage data from Apple iOS keyboards *emojiXpress* [13]. We randomly added five more emojis into the phrase set. The phrase set is provided in Appendix A. We translated the phrases into Chinese for our Chinese participants. Three of the authors translated the phrases and verified the correctness together.

The experimenter sent every phrase to each participant via Voicemoji's chat function. Participants then heard the content of each phrase on their devices, and began to input the phrases using either the Apple iOS keyboard or Voicemoji. They then pressed the *Send* button to send the finished phrase back to the experimenter for verification. After finishing all 27 phrases with one method, participants took a five minute break and then started to input the phrases with the other method.

Emoji Suggestion Session. We sent eight phrases without emojis to our participants, asking them to speak the phrases using

[12]http://www.emojistats.org/
[13]https://www.emojixpress.com/

Voicemoji and to consider the relevance of the emoji suggestions. Specifically, participants were asked to select the emojis that they would use with the spoken phrase from among the five suggestions produced by Voicemoji. The purpose was to evaluate the accuracy and the usefulness of the emoji suggestion feature. Eight phrases were selected from the *Sentiment 140* dataset [22], with four positive sentiment phrases and four negative sentiment phrases. The Sentiment 140 dataset contains 1.6 million Tweets annotated with positive or negative sentiment labels. The phrase set is provided in Appendix B.

Participants were also instructed to compose a phrase by themselves in order to get a sense of using the system in real-settings; therefore, every person evaluated nine phrases in the emoji suggestion session.

After both sessions were complete, participants were asked to rate different methods with the SUS questionnaire [6] and the NASA TLX survey [25], followed by a short debrief on their experiences using our Voicemoji method.

5.4 Metrics

For the emoji entry session, there was one independent variable, *entry method*, with two levels: Voicemoji and the *de facto* Apple iOS keyboard. We measured the entry time from the audio recording, including the *total entry time* and the *emoji entry time*. Because participants were using Apple's built-in VoiceOver screen reader, the audio feedback of all actions could be recorded, allowing us to measure timings from the audio record.

Total entry time started when participants pressed the *Speech* button in the Voicemoji condition, or when they started dictation in the Apple iOS keyboard condition[14]. The timing ended after participants finished each phrase and VoiceOver read it aloud.

Emoji entry time was part of the *total entry time*, and measured only the time for emoji entry. For Voicemoji, it included the time for speaking the commands, playing the results, and choosing the emojis from the results list; for the Apple iOS keyboard, emoji entry time included the time for switching from the alphabetic to the emoji keyboard, and visually searching for and selecting emojis in the list–or, if participants used an emoji shortcut, it included the time for typing the shortcuts and selecting them from the suggestion bar.

Due to network latency and processing time, we removed the waiting time before the results were read by Voiceover in the Voicemoji condition. We did this because we wanted to evaluate the interaction time, rather than the vagaries of Voicemoji's current implementation. Processing time was neither equal nor controlled for the two methods: Voicemoji required the network for API usage and hosting the web page, while the Apple iOS keyboard's functions operated offline. Including the processing time would thus confound the interaction time: the average processing time was 2.2 ± 1.3 seconds for English, and 3.3 ± 1.7 seconds for Chinese; the average network latency was 3.9 ± 1.1 seconds for U.S. participants, and 5.2 ± 3.2 seconds for Chinese participants. The average emoji entry time with/without latency of each participant is presented in Appendix C.

[14] All participants used Apple's so-called "Magic Tap" gesture to start dictation, which was a double tap with two fingers.

The first author went through all of the audio recordings and manually calculated the entry time. These results were then validated by another author, whose results were in agreement.

For the emoji suggestion session, we recorded the number of emojis that were perceived as relevant suggestions. We then calculated the accuracy based on the number of chosen suggestions: *pick-1 accuracy* examined whether any of the five suggested emojis were perceived as relevant. As long as any emoji from the suggestions were picked, the *pick-1 accuracy* was 100% for the suggestion; *overall accuracy* examined how many emojis from the suggestions were perceived relevant on average. For instance, if three of five suggestions were chosen as relevant, then the *overall accuracy* was 60%.

In addition, we also logged the usage of the Voicemoji website during the three-day interval between phase one and phase two of the study. For privacy, we only logged the emoji results and the IP address of the query. The IP address was for differentiating users.

We also analyzed participants' qualitative feedback on Voicemoji. We analyzed the audio transcripts using an open coding method [13]. Two of the authors first translated the Chinese transcripts into English, coded the transcripts individually, and met to achieve consensus on the codes. The codes were then discussed by the research team using affinity diagramming [50]. The codes were created in a random order and iteratively arranged into a hierarchy of themes.

6 QUANTITATIVE RESULTS

In this section, we present the results from our user study. Overall, we collected 648 phrases (27 phrases × 2 methods × 6 participants from each nation × 2 nations) for the *emoji entry task*, and 108 phrases (9 phrases × 6 participants from each nation × 2 nations) for the *emoji suggestion task*.

6.1 Emoji Input with the Apple iOS Keyboard

Before diving into comparative results between the Apple iOS keyboard and Voicemoji, it is useful to characterize emoji input with the Apple iOS keyboard. With this keyboard, participants entered emojis with either the emoji keyboard or the emoji shortcut method, described above. Three American participants only used the emoji keyboard method, three used both methods, and all six Chinese participants used both methods to enter emojis. On average, for the three American participants who used both methods, 13.3 of 27 phrases (49.3%) were completed with the emoji keyboard, and 13.7 of 27 phrases (50.7%) were completed with emoji shortcuts. Chinese participants completed 12.7 of 27 phrases (47.0%) with the emoji keyboard, and 14.3 of 27 phrases (53.0%) with the emoji shortcut.

There were also occasions when participants first tried the shortcut method, but failed to retrieve their desired emoji, and then switched to the emoji keyboard. On average, there were 4.3 of 27 phrases (15.9%) that participants first tried the shortcut method and then switched to the emoji keyboard in the U.S. group, and 5.3 of 27 phrases (19.6%) in the Chinese group.

6.2 Entry Time

Figure 8 shows the *total entry time* and *emoji entry time* of the two methods. The average total entry time for Voicemoji was 6.9 seconds, which was 87.1% shorter than the 56.6 seconds for the Apple iOS keyboard. We log-transformed *total entry time* and *emoji entry time* to make both fit gamma distributions, as is common practice with time measures [37]. We performed analyses of variance using a generalized linear mixed model (GLMM) with gamma link function [39] on *total entry time* and *emoji entry time* separately, treating *entry method* and *nation* as fixed effects, and *participant* and *trial* as random effects. For *total entry time*, we found a significant main effect of *entry method* ($\chi^2_{(1,N=648)} = 486.29, p < .001$), indicating that Voicemoji was significantly faster than the Apple iOS keyboard. The effect of *nation* was not statistically significant ($\chi^2_{(1,N=648)} = 0.09$, *n.s.*). There was a significant interaction between *nation* and *entry method* ($\chi^2_{(1,N=648)} = 9.22, p < .01$), as shown in Figure 9(b).

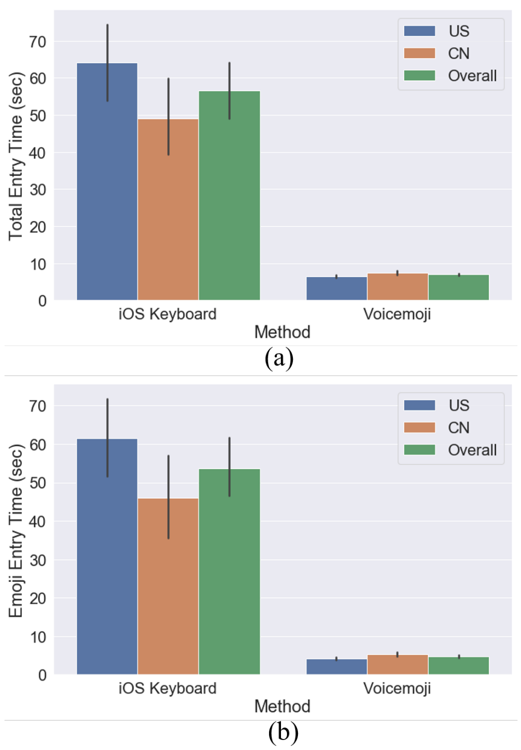

Figure 8: The average (a) *total entry time* and (b) *emoji entry time* of the two methods by nation (they were log-transformed in the analysis). Error bars represent 95% confidence intervals (CIs).

The average *emoji entry time* for Vcoiemoji was 4.7 seconds, which was 91.2% shorter than the 53.7 seconds for the Apple iOS keyboard. We found a significant main effect of *entry method*

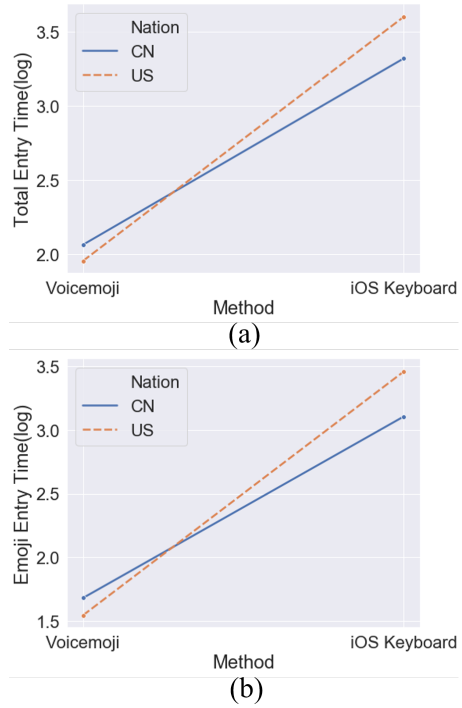

Figure 9: Interaction effect of *Method* × *Nation* on log entry time. Both interaction effects indicate that the time saved by Voicemoji in the U.S. group is more than in the Chinese group.

($\chi^2_{(1,N=648)} = 513.55, p < .001$). There was no statistically significant effect of *nation* ($\chi^2_{(1,N=648)} = 0.06$, *n.s.*). There was also a significant interaction between *nation* and *entry method* ($\chi^2_{(1,N=648)} = 16.11, p < .001$), as shown in Figure 9(b).

From the time distribution (Figure 10), we noticed that the users also entered certain emojis quickly with the iOS keyboard. The could be explained by several observations in the study: 1) participants used certain emojis more frequently (*e.g.*, ❤️) than others (*e.g.*, 😹), so that they were familiar with the keywords of those emojis; 2) when using the emoji keyboard, some participants with residual vision entered the emojis faster as they could utilize the visual information to guide the search procedure.

Taken together, then, our results for *time* make it clear that Voicemoji was significantly faster than the Apple iOS keyboard for entering text with emojis.

6.3 Voicemoji Suggestion Accuracy

The *pick-1 accuracy* for Voicemoji was 100% for each phrase, as all participants picked at least one emoji suggestion for every phrase. The *overall accuracy* was 76.0% for the Chinese group (participants picked 3.80 ± 1.03 emojis as relevant suggestions for each phrase), and 72.2% for the American group (participants picked 3.61± 1.24

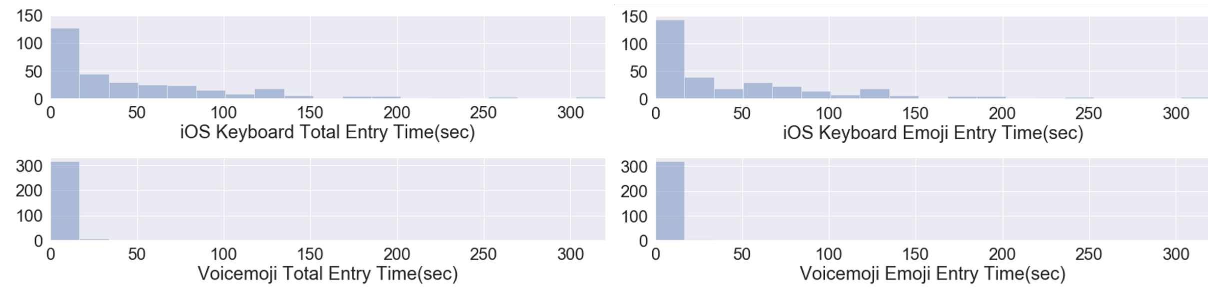

Figure 10: Frequency distribution of the *total entry time* and *emoji entry time* in seconds for the two methods. All the tasks were finished within 20 seconds for Voicemoji, while the distribution were heavy tailed for the Apple iOS keyboard.

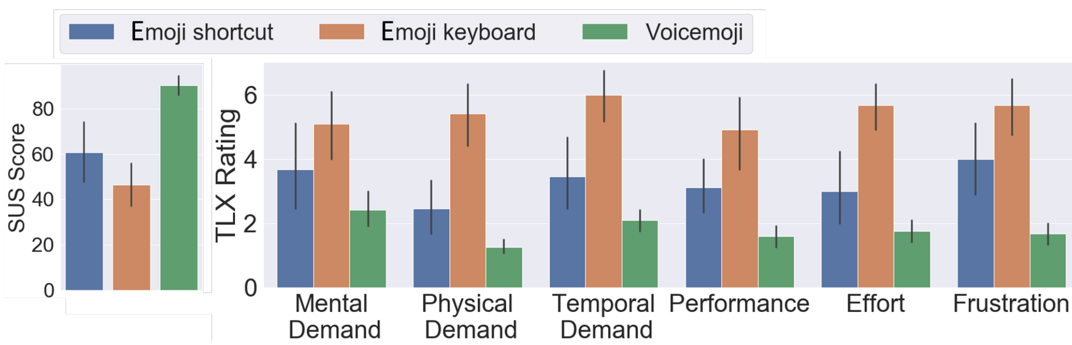

Figure 11: SUS usability scores and NASA TLX workload ratings of different emoji input methods. For SUS scores, higher indicates "more usable" and is better. For NASA TLX ratings, lower indicates "less workload" or "better performance" and is better.

emojis as relevant suggestions for each phrase). These results indicate that the emojis suggested by Voicemoji are generally perceived as relevant by users from both countries.

6.4 Voicemoji Usage

We logged usage data on the Voicemoji website during the three-day interval before the second phase of the study to see whether participants started to use Voicemoji in their daily lives. There were 6 participants (2 from China, 4 from the U.S.) who used the website, resulting in 84 emoji-related commands. On average, participants used the website 4.7 times a day to input emojis. Given that we had told participants that using the Voicemoji tool was optional during this three-day interval, it is hard to draw firm conclusions from this limited usage. However, it does seem that some participants voluntarily began to use Voicemoji in their daily lives even apart from our formal study.

6.5 Subjective Ratings

Our subjective ratings data were ratings on the SUS usability instrument [6] and NASA TLX workload instrument [25], which enabled us to capture usability and workload ratings for both the Apple iOS keyboard and the Voicemoji system. For the Apple iOS keyboard, We separated the emoji keyboard method and the emoji shortcut method when capturing subjective ratings because these methods offered different interaction designs and user experiences. If the participant used both methods in the Apple iOS keyboard condition, then they rated both methods. We therefore had three American participants who only rated the emoji keyboard and Voicemoji, while the other nine participants rated all three methods (i.e., the emoji keyboard, emoji shorcuts, and Voicemoji). Both SUS and TLX results are graphed in Figure 11. We calculated the SUS score on a scale of 0-100, where higher indicates "more usable" [15]. The average SUS score was 90.4 (*SD*=8.11) for Voicemoji, 46.5 (*SD*=16.8) for the emoji keyboard, and 60.8 (*SD*=21.5) for emoji shortcuts. Because of network latency issues, some participants mentioned lowering their SUS scores for Voicemoji.

We performed the nonparametric Aligned Rank Transform procedure [67] on the SUS scores and NASA TLX ratings to examine any effects of *Method*. We found that for SUS scores, *Method* had a statisitcally significant effect ($F_{2,32} = 27.70, p < .001$). Pairwise comparisons with the Bonferroni correction showed that the Voicemoji received higher scores than the emoji keyboard ($p < .01$) and emoji shortcuts ($p < .05$), while there was no significant difference between the emoji keyboard and emoji shortcuts. For NASA TLX ratings, we found that *Method* had a statistically significant effect on all dimensions: *mental demand* ($F_{2,32} = 8.35, p < .005$), *physical demand* ($F_{2,32} = 87.80, p < .001$), *temporal demand* ($F_{2,32} = 26.95, p < .001$), *performance* ($F_{2,32} = 16.64, p < .001$), *effort* ($F_{2,32} = 30.49, p < .001$) and *frustration* ($F_{2,32} = 27.13, p < .001$).

[15]https://measuringu.com/sus/

Pairwise comparisons with the Bonferroni correction showed that Voicemoji had significantly lower *physical demand* and *frustration* than emoji shortcuts ($p < .05$); and had significantly lower (better) scores on all ratings than the emoji keyboard ($p < .05$).

7 QUALITATIVE RESULTS

Our affinity diagramming analysis revealed four main themes arising in participants' feedback about Voicemoji: *conveniently entering emojis*, *helpful suggestions*, *supporting emoji exploration*, and *enriching communication*. We describe each of these in turn.

Conveniently entering emojis. All participants mentioned that finding and entering emojis with Voicemoji was much easier than with the Apple iOS keyboard. As most participants were already familiar with speech input generally, they quickly mastered Voicemoji after learning the commands. Participants consistently mentioned two benefits of Voicemoji: (1) *supporting both fuzzy and precise search*, and (2) *the ease of speech input*. Four participants appreciated the flexibility with which Voicemoji queries could be formulated, namely that they could either speak the name of the emoji, or describe emoji when unsure of their names. As P6 said, "It is good that Voicemoji can understand my description. I can use [the] 'emoji search' command if I'm not confident, and use the 'insert' command when I know the exact name, which is very efficient." Three participants mentioned feeling it was great that they could input emojis with voice. As P4 said, "[Voicemoji is] very convenient when I want to add some emojis when I'm speaking a long paragraph, I can just do it at once without switching the keyboard."

Helpful suggestions. Four participants mentioned that Voicemoji's suggestion feature was useful in helping them find the right emojis. As P12 said, "I think it was really helpful for just even quickly composing messages that have the right emoji in them, and there's really nothing like Voicemoji's suggestion function." Participant 2 also said, "some suggestions were even better than my intended emojis." Participants 1 and 6 mentioned that even when they had no intention of adding an emoji, the emoji suggestions provided them with appropriate emojis, making it easy to just tap one to insert it.

Supporting emoji exploration. Four participants mentioned that Voicemoji provided a new way to learn new emojis. The multiple options provided by Voicemoji sometimes exposed emojis that were previously unfamiliar to users. As P1 said, "It will suggest some emojis I've never imagined, such as an emoji about a particular object." Participant 2 agreed, saying that as a result, the unexpected emojis "add surprise and delight of using emojis, and offer a new way to learn them." Participant 2 also mentioned that finding rare emojis was a painful process in the Apple iOS keyboard, while Voicemoji's entry process was equally efficient for all emojis.

Enriching communication. Participants 6 and 8 explicitly talked about Voicemoji enriching their daily communications. Participant 6 did not use emojis much before the study, but he found that using Voicemoji lowered the effort of finding emojis, and he started to send emojis during his everyday messaging, which "improved the expressiveness and the interactivity." Participant 8 reported using emojis frequently
before our study, and she found that using Voicemoji made the process even easier and faster, making her "feel as mainstream as my sighted friends."

8 DISCUSSION

In this work, we have presented *Voicemoji*, a speech-based emoji input method. In creating Voicemoji, we sought to address four major challenges of existing emoji entry methods, namely that they are *time consuming*, *inconsistent with users' expectations*, *lacking support for discovering emojis*, and *lacking support for finding the right emojis*. To address these issues, we designed several features including speech-based search and automatic emoji suggestions. Our user study showed that finding and entering emojis with Voicemoji was significantly faster than the current Apple iOS keyboard, and most of the suggested emojis were perceived as relevant to users' spoken content. Importantly, Voicemoji's apparent benefits were consistent for both our American and Chinese participants, suggesting that the design of Voicemoji is at least somewhat generalizable across different languages.

Many participants expressed that they would use the Voicemoji even after the study. Participant 10 commented that he was doubtful about the usability at first, but after the study, he commented that Voicemoji was, "practical as hell" and he "would immediately use the tool" in his everyday life. The feedback given by P6 and P8 about Voicemoji "enriching communication" indicated that the impact of Voicemoji was not only on the speed or effort of entering emojis, but also on a subjectively higher level, where Voicemoji provided support for people to express themselves better. Emojis improve expressiveness, and technologies like Voicemoji that facilitate their entry can improve this expressiveness.

One piece of feedback we did not anticipate was that participants mentioned the potential usefulness of Voicemoji beyond the blind community. Both P3 and P8 mentioned that the design of speaking emojis could actually "benefit beyond blind people, as many of us use voice input sometimes" (P3). Adopting a perspective of ability-based design [68, 69], technologies designed for users with disabilities are often more usable to wide range of users [43, 60]. For example, when people are walking or driving, they might benefit from using speech more because of so-called "situational impairments" [66], and Voicemoji could facilitate emoji entry in such situations.

8.1 Limitations

As with any research project, there are several limitations of this work. First, we only conducted studies with existing Apple iOS users because of the consistency of the system. Different Android phones have different voice-over descriptions for the same emojis, and not all emojis have descriptions on Android systems compared to iOS. However, some Android keyboards such as *Gboard* provide a built-in emoji search function, which might result in better performance than the Apple iOS keyboard. Second, our study sessions were all conducted remotely, making them somewhat less controlled than a typical lab-based study. For example, participants used their own Apple iPhone models, which could be somewhat different from one another. Network latency was also not controlled during our study, where a lab-based study on our university campus could achieve shorter response times from the server. However, as the purpose of the comparison is to evaluate the timing of the two interactions (*i.e.* the search interaction, the command design, the voice feedback), rather than the current implementation performance, we did not include the latency in our analysis. Third, the Google speech-to-text API we used would cut off the speech after a pause longer than two seconds, which meant that participants could not pause too long during their speech. Also, Voicemoji did not support real-time recognition due to network issues, which hindered the user experience. Fourth, although we logged website usage during the three-day interval between study phases, conducting a long-term field deployment would reveal insights on how participants use Voicemoji in their daily communications, if at all. Our preliminary evidence suggests that some participants would indeed use Voicemoji in their daily lives, but this result is anecdotal at this point.

9 FUTURE WORK

Beyond remedying or addressing the limitations described above, there are exciting directions for future work based on this initial study. Six participants mentioned that it was cumbersome to switch between apps to use Voicemoji, as it was a web app, and so a priority would be to build Voicemoji into an actual keyboard, which is supported on Android devices. We already have open-sourced our implementation of Voicemoji, and keyboard developers could consider adding it to their keyboard projects. We also see value in adding explanations and example uses emojis. For example, P6 expressed that sometimes he could not understand an emoji just by its name, and many emojis have similar names. Therefore, it would be helpful if Voicemoji could also provide explanations of how to use emojis. Finally, we also think the style of interaction employed by Voicemoji could be extended to other forms of visual media besides emojis, such as stickers and memes [20].

10 CONCLUSION

In this work, we have presented *Voicemoji*, a speech-based emoji input method to make emoji input, search, and discovery easier and more accessible for blind or low vision (BLV) users. We conducted an interview study to understand the current emoji entry experience and challenges for BLV users, and designed multiple novel Voicemoji features, including direct and fuzzy emoji search, speech-only emoji insertion, color or skin tone emoji modification, and automatic emoji suggestions to address the challenges. Results from our formal evaluation with both American and Chinese participants demonstrated that Voicemoji provided significantly faster entry times, greater perceived usability, and lower perceived workload for both groups of participants than the current Apple iOS keyboard. We hope that by open-sourcing our implementation, Voicemoji will encourage keyboard developers to consider designing inclusive input methods for emojis. We also hope that this work will inspire future research in non-textual information entry methods.

ACKNOWLEDGMENTS

The authors thank Yiming Zhang, Fanwen Ji, Yuying Liu for participating the preliminary design and evaluation process. The work was supported in part by Baidu Inc. Any opinions, findings, conclusions or recommendations expressed in this work are those of the authors and do not necessarily reflect of any supporter.

REFERENCES

[1] Kim Albarella. 2018. The Secret Language of Emoji. https://staysafeonline.org/blog/secret-language-emoji/
[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz. 2019. Guidelines for Human-AI Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 113. https://doi.org/10.1145/3290605.3300233
[3] 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* (Bellevue, Washington) *(ASSETS '13)*. Association for Computing Machinery, New York, NY, USA, Article 11, 8 pages. https://doi.org/10.1145/2513383.2513440
[4] Francesco Barbieri and Jose Camacho-Collados. 2018. How Gender and Skin Tone Modifiers Affect Emoji Semantics in Twitter. In *Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics*. Association for Computational Linguistics, New Orleans, Louisiana, 101–106.
[5] Nicole Bleuel. 2016. Promoting gender equality through emoji. https://blog.google/products/android/promoting-gender-equality-through-emoji/
[6] John Brooke. 1996. SUS: a quick and dirty usability scale. *Usability evaluation in industry* 1 (1996), 189.
[7] Charlotte Buff. 2017. Proposal for Fully Gender-Inclusive Emoji.
[8] Jeremy Burge. 2017. 5 Billion Emojis Sent Daily on Messenger. https://blog.emojipedia.org/5-billion-emojis-sent-daily-on-messenger/
[9] Jeremy Burge. 2020. On Families and Equality. https://blog.emojipedia.org/on-families-and-equality/
[10] Zhenpeng Chen, Xuan Lu, Sheng Shen, Wei Ai, Xuanzhe Liu, and Qiaozhu Mei. 2017. Through a Gender Lens: An Empirical Study of Emoji Usage over Large-Scale Android Users.
[11] Yuri Choi, Kyung Hoon Hyun, and Ji-Hyun Lee. 2020. Image-Based Tactile Emojis: Improved Interpretation of Message Intention and Subtle Nuance for Visually Impaired Individuals. *HumanComputer Interaction* 35, 1 (2020), 40–69.
[12] Neil Cohn, Jan Engelen, and Joost Schilperoord. 2019. The grammar of emoji? Constraints on communicative pictorial sequencing.
[13] Juliet Corbin and Anselm Strauss. 2014. Basics of qualitative research: Techniques and procedures for developing grounded theory.
[14] Sam Cox. 2019. World Emoji Day: making emojis more accessible. https://wearesocial.com/uk/blog/2019/07/world-emoji-day-making-emojis-more-accessible
[15] Henriette Cramer, Paloma de Juan, and Joel Tetreault. 2016. Sender-Intended Functions of Emojis in US Messaging. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 504509.
[16] Mark Davis and Peter Edberg. 2020. Unicode Emoji. https://unicode.org/reports/tr51/
[17] Jan Derboven, Jonathan Huyghe, and Dirk De Grooff. 2014. Designing Voice Interaction for People with Physical and Speech Impairments. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational* (Helsinki, Finland) *(NordiCHI '14)*. Association for Computing Machinery, New York, NY, USA, 217226.
[18] Emojipedia. 2020. Emoji Version 13.0 List. https://emojipedia.org/emoji-13.0/
[19] Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm.
[20] Cole Gleason, Amy Pavel, Xingyu Liu, Patrick Carrington, Lydia B. Chilton, and Jeffrey P. Bigham. 2019. Making Memes Accessible. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility* (Pittsburgh, PA, USA) *(ASSETS '19)*. Association for Computing Machinery, New York, NY, USA, 367376. https://doi.org/10.1145/3308561.3353792
[21] Cole Gleason, Amy Pavel, Emma McCamey, Christina Low, Patrick Carrington, Kris M. Kitani, and Jeffrey P. Bigham. 2020. Twitter A11y: A Browser Extension to Make Twitter Images Accessible. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 112. https://doi.org/10.1145/3313831.3376728
[22] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision.
[23] Susumu Harada, Jacob O. Wobbrock, and James A. Landay. 2007. Voicedraw: A Hands-Free Voice-Driven Drawing Application for People with Motor Impairments. In *Proceedings of the 9th International ACM SIGACCESS Conference on*

Computers and Accessibility (Tempe, Arizona, USA) *(Assets '07)*. Association for Computing Machinery, New York, NY, USA, 2734.

[24] Susumu Harada, Jacob O. Wobbrock, Jonathan Malkin, Jeff A. Bilmes, and James A. Landay. 2009. Longitudinal Study of People Learning to Use Continuous Voice-Based Cursor Control. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 347356.

[25] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology* 52 (1988), 139 – 183.

[26] A. G. Hauptmann. 1989. Speech and Gestures for Graphic Image Manipulation. *SIGCHI Bull.* 20, SI (March 1989), 241245.

[27] Susan C. Herring and Ashley R. Dainas. 2020. Gender and Age Influences on Interpretation of Emoji Functions. *Trans. Soc. Comput.* 3, 2, Article 10 (April 2020), 26 pages. https://doi.org/10.1145/3375629

[28] Jiaxiong Hu, Qianyao Xu, Limin Paul Fu, and Yingqing Xu. 2019. Emojilization: An Automated Method For Speech to Emoji-Labeled Text. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 16. https://doi.org/10.1145/3290607.3313071

[29] Albert H. Huang, David C. Yen, and Xiaoni Zhang. 2008. Exploring the Potential Effects of Emoticons. *Inf. Manage.* 45, 7 (Nov. 2008), 466473.

[30] Takeo Igarashi and John F. Hughes. 2001. Voice as Sound: Using Non-Verbal Voice Input for Interactive Control. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology* (Orlando, Florida) *(UIST '01)*. Association for Computing Machinery, New York, NY, USA, 155156.

[31] Apple Inc. 2020. iOS Dictation. https://support.apple.com/en-us/HT208343

[32] Minal Jain, Sarita Seshagiri, and Simran Chopra. 2016. How Do I Communicate My Emotions on SNS and IMs?. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 767774.

[33] Shao kang Lo. 2008. The Nonverbal Communication Functions of Emoticons in Computer-Mediated Communication. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* 11 5 (2008), 595–7.

[34] Ryan Kelly and Leon Watts. 2015. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. Experiences of Technology Appropriation: Unanticipated Users, Usage, Circumstances, and Design ; Conference date: 20-09-2015 Through 20-09-2015.

[35] Sujay Khandekar, Joseph Higg, Yuanzhe Bian, Chae Won Ryu, Jerry O. Talton Iii, and Ranjitha Kumar. 2019. Opico: A Study of Emoji-First Communication InAMobileSocialApp. In *Companion Proceedings of The 2019 World Wide Web Conference* (San Francisco, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 450458. https://doi.org/10.1145/3308560.3316547

[36] Joon-Gyum Kim, Taesik Gong, Evey Huang, Juho Kim, Sung-Ju Lee, Bogoan Kim, JaeYeon Park, Woojeong Kim, Kyungsik Han, and JeongGil Ko. 2019. Bringing Context into Emoji Recommendations (Poster). In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services* (Seoul, Republic of Korea) *(MobiSys '19)*. Association for Computing Machinery, New York, NY, USA, 514515. https://doi.org/10.1145/3307334.3328601

[37] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. 2001. Log-normal Distributionsacross the Sciences: Keys and Clues. *BioScience* 51, 5 (05 2001), 341–352. https://doi.org/10.1641/0006-3568(2001)051[0341:LNDATS]2.0.CO;2 arXiv:https://academic.oup.com/bioscience/article-pdf/51/5/341/26891292/51-5-341.pdf

[38] Haley MacLeod, Cynthia L. Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 59885999. https://doi.org/10.1145/3025453.3025814

[39] Charles E. McCulloch and John M. Neuhaus. 2005. *Generalized Linear Mixed Models.* American Cancer Society, New York, Chapter 1, 1. https://doi.org/10.1002/0470011815.b2a10021

[40] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and Inter-Rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 72 (Nov. 2019), 23 pages.

[41] MDN contributors. 2020. MDN Web Docs - ARIA. https://developer.mozilla.org/en-US/docs/Web/Accessibility/ARIA

[42] Meredith Ringel Morris, Annuska Zolyomi, Catherine Yao, Sina Bahram, Jeffrey P. Bigham, and Shaun K. Kane. 2016. "With Most of It Being Pictures Now, I Rarely Use It": Understanding Twitter's Evolving Accessibility to Blind Users. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 55065516. https://doi.org/10.1145/2858036.2858116

[43] Martez E. Mott, Radu-Daniel Vatavu, Shaun K. Kane, and Jacob O. Wobbrock. 2016. Smart Touch: Improving Touch Accuracy for People with Motor Impairments with Template Matching. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 19341946. https://doi.org/10.1145/2858036.2858390

[44] Pandu Nayak. 2019. Understanding searches better than ever before. https://blog.google/products/search/search-language-understanding-bert/

[45] Daniel S. Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D. Cubuk, and Quoc V. Le. 2019. SpecAugment: A Simple Data Augmentation Method for Automatic Speech Recognition. In *INTERSPEECH*. IEEE, Austria, 10.

[46] Henning Pohl, Christian Domin, and Michael Rohs. 2017. Beyond Just Text: Semantic Emoji Similarity Modeling to Support Expressive Communication 👫📲😃. *ACM Trans. Comput.-Hum. Interact.* 24, 1, Article 6 (March 2017), 42 pages.

[47] K R Prajwal, C V Jawahar, and Ponnurangam Kumaraguru. 2019. Towards Increased Accessibility of Meme Images with the Help of Rich Face Emotion Captions. In *Proceedings of the 27th ACM International Conference on Multimedia* (Nice, France) *(MM '19)*. Association for Computing Machinery, New York, NY, USA, 202210. https://doi.org/10.1145/3343031.3350939

[48] David Rodrigues, Diniz Lopes, Mar ı́ lia Prada, Dominic Thompson, and Margarida V. Garrido. 2017. A frown emoji can be worth a thousand words: Perceptions of emoji use in text messages exchanged between romantic partners. *Telematics and Informatics* 34, 8 (2017), 1532 – 1543.

[49] Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James A. Landay. 2018. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 159 (Jan. 2018), 23 pages.

[50] RAYMOND SCUPIN. 1997. The KJ Method: A Technique for Analyzing Data Derived from Japanese Ethnology. *Human Organization* 56, 2 (1997), 233–237.

[51] Ashley Sheridan. 2019. Emoji and Accessibility. http://www.ashleysheridan.co.uk/blog/Emoji+and+Accessibility

[52] Xavier Snelgrove. 2016. Teaching Robots to Feel: Emoji & Deep Learning. https://getdango.com/emoji-and-deep-learning/

[53] SwiftKey. 2015. SwiftKey Emoji Report. https://www.scribd.com/doc/262594751/SwiftKey-Emoji-Report

[54] Garreth W. Tigwell and David R. Flatla. 2016. Oh That's What You Meant! Reducing Emoji Misunderstanding. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 859866.

[55] Garreth W. Tigwell, Benjamin M. Gorman, and Rachel Menzies. 2020. Emoji Accessibility for Visually Impaired People. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 114.

[56] Tornadoweb. 2020. Tornado Web Server. https://www.tornadoweb.org/en/stable/

[57] Veroniiiica. 2018. How Do People with Vision Impairments . . . Use Emoji? https://www.perkinselearning.org/technology/blog/how-do-people-vision-impairments-use-emoji

[58] Keith Vertanen and Per Ola Kristensson. 2011. A Versatile Dataset for Text Entry Evaluations Based on Genuine Mobile Emails. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services* (Stockholm, Sweden) *(MobileHCI '11)*. Association for Computing Machinery, New York, NY, USA, 295298. https://doi.org/10.1145/2037373.2037418

[59] Joseph B. Walther and Kyle P. D'Addario. 2001. The Impacts of Emoticons on Message Interpretation in Computer-Mediated Communication. *Social Science Computer Review* 19 (2001), 324–347.

[60] Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, and Yuanchun Shi. 2019. EarTouch: Facilitating Smartphone Use for Visually Impaired People in Mobile and Public Scenarios. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 113. https://doi.org/10.1145/3290605.3300254

[61] L é onie Watson. 2016. Accessible emoji. https://tink.uk/accessible-emoji/

[62] WHIRLSCAPE INC. 2016. Dango - Your Emoji Assistant. https://getdango.com

[63] Sanjaya Wijeratne, Lakshika Balasuriya, Amit Sheth, and Derek Doran. 2017. A Semantics-Based Measure of Emoji Similarity. In *Proceedings of the International Conference on Web Intelligence* (Leipzig, Germany) *(WI '17)*. Association for Computing Machinery, New York, NY, USA, 646653.

[64] Wikipedia. 2020. Gboard - a virtual keyboard app developed by Google. https://en.wikipedia.org/wiki/Gboard

[65] Sarah Wiseman and Sandy J. J. Gould. 2018. Repurposing Emoji for Personalised Communication: Why 🍕 means "I Love You". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 110.

[66] Jacob O. Wobbrock. 2019. Situationally Aware Mobile Devices for Overcoming Situational Impairments. In *Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (Valencia, Spain) *(EICS '19)*. Association for Computing Machinery, New York, NY, USA, Article 1, 18 pages.

https://doi.org/10.1145/3319499.3330292
[67] Jacob O. Wobbrock, Leah Findlater, Darren Gergle, and James J. Higgins. 2011. The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only Anova Procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 143146. https://doi.org/10.1145/1978942.1978963
[68] Jacob O. Wobbrock, Krzysztof Z. Gajos, Shaun K. Kane, and Gregg C. Vanderheiden. 2018. Ability-Based Design. *Commun. ACM* 61, 6 (May 2018), 6271.
[69] Jacob O. Wobbrock, Shaun K. Kane, Krzysztof Z. Gajos, Susumu Harada, and Jon Froehlich. 2011. Ability-Based Design: Concept, Principles and Examples. *ACM Trans. Access. Comput.* 3, 3, Article 9 (April 2011), 27 pages. https://doi.org/10.1145/1952383.1952384
[70] Rui Zhou, Jasmine Hentschel, and Neha Kumar. 2017. Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 748759. https://doi.org/10.1145/3025453.3025800

Appendices

A PHRASE SET FOR THE EMOJI ENTRY SESSION

	Emoji Description	English Phrases	Chinese Phrases
1	Hamburger	Are you going to join us for lunch🍔	你要不要和我们一起吃午饭🍔
2	Folded hands	Keep me posted please! 🙏	有新消息请通知我🙏
3	Crying face	I am so sorry 😢	真的很对不起😢
4	Sun	Florida is great ☀️	海南省很漂亮☀️
5	Ear	What do you hear 👂	你听见什么了👂
6	Wrapped gift	A 🎁 isn't necessary	没必要送我🎁
7	Musical note	He would love anything about rock 🎵	只要是摇滚🎵他都喜欢
8	School	And leave me at 🏫 alone	把我留在🏫就行
9	Airplane	I'm on a ✈️	我在✈️上
10	Hundred points	I think that answer is 💯	我觉得那个答案💯
11	Smiling face with open mouth	😃 How are you	😃 最近怎么样
12	Raised fist	✊ I am trying again	✊ 我再试一次
13	Disappointed face	😞 We are all fragile	😞 人都是脆弱的
14	Eyes	👀 See you soon	👀改天见
15	OK hand sign	👌I'll get you one.	👌我帮你　一个
16	Face with tears of joy	😂	😂
17	Red heart	❤️	❤️
18	Smiling face with heart-shaped eyes	😍	😍
19	Loudly crying face	😭	😭
20	Thumb up	👍	👍
21	Face with hand over mouth	🤭	🤭
22	Face with rolling eyes	🙄	🙄
23	Probing cane	🦯	🦯
24	Cat face with tears of joy	😹	😹
25	Woman gesturing no	🙅‍♀️	🙅‍♀️

B PHRASE SET FOR THE EMOJI SUGGESTION SESSION

	Sentiment	English Phrases	Chinese Phrases
1	Negative	Can't sleep. my tooth is aching.	我的牙齿痛得睡不着觉了
2	Negative	So tired, I didn't sleep well at all last night.	太累了，昨晚没睡好
3	Negative	Math review, I'm going to fail the exam.	我在复习数学考试，感觉要不及格了
4	Negative	Still hungry after eating	吃完以后还是很饿
5	Positive	My dentist appointment today was quite enjoyable.	今天去医院，检查的结果很好
6	Positive	I'm moving to a new place	我就要搬家啦
7	Positive	Just listened to a new song and It was good	刚刚听了一首很好听的歌
8	Positive	Happy birthday to you	祝你生日快乐

C MEAN EMOJI ENTRY TIME WITH/WITHOUT DELAY

Participant ID	Voicemoji (without delay)	Voicemoji (with delay)	iOS Keyboard
P1	3.2s	7.7s	49.4s
P2	6.4s	8.4s	42.0s
P3	3.8s	7.3s	50.4s
P4	5.6s	13.6s	64.2s
P5	5.9s	9.9s	27.5s
P6	6.6s	12.1s	42.4s
P7	3.3s	5.3s	21.0s
P8	5.1s	9.6s	102.8s
P9	5.1s	10.1s	105.8s
P10	4.0s	7.5s	46.7s
P11	4.6s	8.6s	63.2s
P12	2.8s	4.8s	29.1s