"They Like to Hear My Voice": Exploring Usage Behavior in Speech-Based Mobile Instant Messaging

GABRIEL HAAS, Ulm University, Germany
JAN GUGENHEIMER, Telecom-Paris/LTCI/IP-Paris, France
JAN OLE RIXEN, Ulm University, Germany
FLORIAN SCHAUB, University of Michigan, USA
ENRICO RUKZIO, Ulm University, Germany

Adoption and use of smartphone-based asynchronous voice messaging has increased substantially in recent years. However, this communication channel has a strong tendency to polarize. To provide an understanding of this modality, we started by conducting an online survey (n=1,003) exploring who is using voice messages, their motives, and utilization. In a consecutive field study (n=6), we analyzed voice messaging behavior of six avid voice message users in a two-week field study, followed by semi-structured interviews further exploring themes uncovered in our survey. Conducting a thematic analysis, we identified four themes driving voice messaging usage: *convenience*, *para-linguistic* features, *situational* constraints and the *receiver*. Voice messaging helps to overcome issues of mobile communication, through ease of use, asynchronous implementation, and voices' rich emotional context. It also was perceived as enabling more efficient communication, helps to handle secondary occupations, and better facilitates maintenance of close relationships. Despite the increased effort required to listen to a voice message, they complement communication with people we care about.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**; *Empirical studies in collaborative and social computing*.

Additional Key Words and Phrases: mobile instant messaging; voice messaging; messenger applications

ACM Reference Format:
Gabriel Haas, Jan Gugenheimer, Jan Ole Rixen, Florian Schaub, and Enrico Rukzio. 2020. "They Like to Hear My Voice": Exploring Usage Behavior in Speech-Based Mobile Instant Messaging. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20), October 5–8, 2020, Oldenburg, Germany*. ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3379503.3403561

1 INTRODUCTION

Nowadays, interpersonal communication often occurs via internet-connected mobile devices. During the last decade, more traditional communication channels such as calling and texting, have been extended by mobile instant messaging apps. Those messenger apps allow the sharing of rich media such as pictures, videos and every other type of media. Only recently a new communication channel arose: voice messaging. Voice messaging usage is growing rapidly and reached 13 voice messages (VMs) per user and day on WhatsApp in 2014 [28] or 6.1 billion VMs per day on WeChat in 2017 [34].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2020 Copyright held by the owner/author(s). Publication rights licensed to ACM.
Manuscript submitted to ACM

In theory, voice messaging combines two advantages from text messaging and calling. Due to its asynchronous implementation, it does not require simultaneous availability of sender and receiver but retains the freedom to answer when it suits the person's context, similar to mobile text messaging. Other than the traditional voice mail system, the mobile implementation that is integrated into the chat window is much more immediate and not lagging behind. It allows for fast responses and creates a dialog flow that is similar to text based chatting. Furthermore by using voice as the medium, users are able to express themselves more naturally and transmit more emotions than in text [6].

However, we do not exactly know how this new communication channel is utilized in practice, what are users' rationales, and especially how well does voice messaging handle the always changing context of mobile users. Insights into this usage behavior will help to understand the benefits voice messaging provides and how it can be utilized more effectively (e.g. providing voice transcripts in situations in which audio playback is unsuitable or hardly possible). The goal of this work is to build an understanding of the WHY and HOW of voice message use and especially the users who choose to use this polarizing means of communication extensively.

To be able to provide insights into voice messaging usage, we conducted an online survey with 1,003 participants to understand who is using voice messaging and for what reasons. We collected participants usage behavior, key demographics, and motivation of use as open text responses. The free text responses were analyzed using thematic analysis and uncovered four main themes of motivation. To validate and strengthen the found themes, we further conducted a two-week field study with six avid voice messaging users. In a final step, the quantitative findings (logged usage behavior) and qualitative findings (interviews) of the field study were used to further explain and support our four main themes of motivation.

Based on our online survey, we found that voice messaging has a wide and diverse user base and describe differences between casual and daily users (shifted towards younger male users). We identified four major themes that categorize motivations for voice messaging: *convenience*, *para-linguistic* features, *situational* constraints, and *receiver* preferences or handicaps. During our field study, we logged all events of participant's received or sent VMs and corresponding meta-information (846 events in total), allowing us to analyze frequency and nature of use. Using experience sampling and activity recognition, we included measures of usage context and sender-receiver relationship. We found that almost half of the voice message interactions were performed while the user was in motion such as walking, driving, bicycle riding or in transitions between contexts. Participants reported high perceived nearness to receiver/sender of messages; and voice messaging was perceived as joyful. Interviews provided further insights such as retrieval problems and an shift of effort towards the receiver of a message.

2 RELATED WORK

Our work is based on studies of (mobile) instant messaging in its many manifestations in the field of computer-mediated communication (CMC). CMC encompasses any human communication that is performed with the help of two or more electronic devices [21]. The effects of instant messaging and digital social networks on individuals have been a common topic in research. Especially teenagers are often the subject of investigation since mobile messaging has become the primary way they communicate with friends [17]. Aspects that have been evaluated range from impacts on language [14] and challenges for parenting [37] to video chatting [3].

2.1 Voice Messaging

Voice messaging is a specific modality of mobile instant messaging. Especially for younger users, asynchronous communication is often preferred over conventional telephony [17], as it does not require the simultaneous availability

of the involved parties. In addition, VMs can be recorded faster than typing a text message [30], allow for greater expressiveness [6] and deeper emotional bonding [32].

However, VMs also have drawbacks compared to text messaging. First, speech is inherently public. The sender is affected during recording, as everybody around them is able to hear their voice. The receiver may be similarly affected during playback of a message, unless they are using headphones or the phones ear-speaker. This characteristic of VMs could be an impediment to adoption and use due to privacy concerns, as already found during use of personal voice assistants (e.g., Siri) in public [23]. Second, sound is transient which makes it cumbersome to review and edit [33]. If a slip-up occurs during recording, the whole message has to be re-recorded. On the receiver side, messages are tedious to scan [33]. For example, finding an appointment's date and location conveyed in a voice message, requires listening to the whole message or browsing the message to find the right point in time.

2.2 Texting versus Calling

In the context of mobile cellphone usage, Rettie, and Reid et al. identified distinct user groups that prefer either to talk on the phone or to text [26, 27]. Talkers, who prefer to talk on the phone but use text messages as a convenient complementary medium, and texters who feel uncomfortable on the phone and prefer to send text messages. Rettie, and Reid et al. explain the distinction between the two groups and their aversion to telephony, which they attribute to difficulties in self-presentation and personality traits. Therefore, we also captured and investigated the personality traits of users of voice messaging.

2.3 Messenger Apps

Since most mobile phone contracts nowadays include data plans, mobile messaging has shifted from SMS to messenger apps. Church et al. [8] looked into how messaging app usage differs from SMS practices. Technical differences are that they allow the exchange of almost any kind of media, including pictures, videos and location. Therefore, they create a more social and informal conversation. As a result, since they are free of cost, more messages are being sent. They also found that SMS is perceived as being more reliable and privacy preserving.

Nouwens et al. [25] investigated differences among messenger apps. Many persons use several platforms for communication, conversations and friends are divided among them. Reasons they identified as explaining this division are the behavior of contacts, dynamics of their relationships, and the limitations of technology, but also perceived goals and emotional connotations of messenger apps. Buschek et al. [4] tried to address the lack of expressiveness and emotional awareness of text messaging by augmenting chats with information about users and contexts. They present three different approaches and discuss the lessons learned from deploying those to users. Although they give a detailed summary about chat augmentations, their scope did not include language as a supporting modality, which, according to our findings, could also address those issues in text messaging.

2.4 Characteristics of Communication Channels

Besides differences in communication technology, researchers have also investigated the distinct characteristics of communication channels such as meeting in-person, video chatting, calling, and texting. In software-based communication systems, especially for historical reasons, text-based communication had been used frequently. Only when cameras, network bandwidth, and the required computing power became widely available, audio and video chat became an option. Sherman et al. [32] compared the mentioned conditions with regard to bonding between friends. They observed the largest effect on bonding during personal interaction, followed by video chat, audio chat and IM, in the given order.

Demographics	
N	1,003
Age	18 - 84 (Mdn=32, SD = 11.845)
Gender	44.9% male
	54.7% female
	.4% non-binary
Education	2.6% PhD
	56.2% bachelor/master
	16.7% associate degree
	24.7% high school
	.4% none
Occupation	59.2% full-time employed (>39 h/week)
	25.3% part-time employed (<39 h/week)
	11.0% unemployed
	4.5% other

Table 1. The table provides an overview to the demographic data of our online survey participants.

Despite the use of textual association references such as emoticons, typified laughter and excessive capitalization of letters, IM was found to have a significantly lower degree of bonding than audio chat.

2.5 Channel Choice

El-Shinnawy et al. investigated in 1997 how electronic mail compares to voice mail [11]. They found that the behavior of their participants was not well explained by media richness theory but other factors, more specific to new types of media. This aligns with our findings which indicate that ease of use is one of the most important factors for voice messaging adoption. The selection of communication channels has also been studied in the context of romantic couples. Scissors and Gergle [31], focusing on conflicts, found that couples changed channels for very specific reasons such as conflict escalation, dealing with their own emotions and attempting to resolve the conflict. We found that such reasons also exist for switching from texting to voice messaging. Cramer and Jacobs [10] explored how different channels support couples' needs in a broader context. Based on interviews with 10 couples, they outlined the used channels and the triggers for in-couple communication. Furthermore, they describe strategies of channel switching such as reinforcement by reminding. They also identified channel switching as a strategy to deal with contextual changes which is present as a user rationale in voice messaging as well.

3 ONLINE SURVEY

To develop an understanding of how voice messaging is distributed among smartphone users and what are the factors driving it, we created an online survey.

3.1 Procedure

Once participants accepted our MTurk HIT, they were directed to a self-hosted online survey. After survey completion, participants were shown a unique completion code for the HIT. Median completion time were 3.34 minutes and participants received 0.20 USD for compensation. The survey consisted of three parts. First, we asked about usage frequency and behavior regarding text and voice messaging with messenger applications. We included an open-response question asking for reasons why voice messaging is used over text messaging. The second part was comprised of the

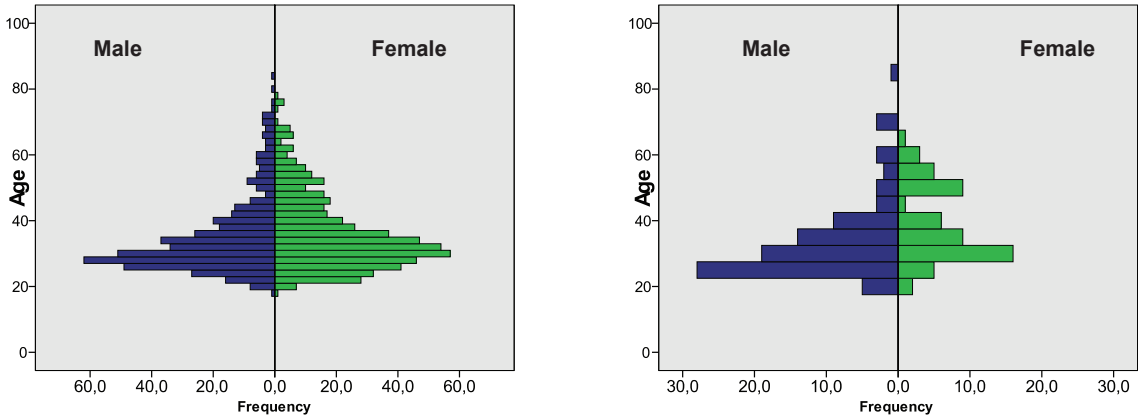

Fig. 1. Left: The distribution of age and gender of our study population (excluding non-binary gender). Right: Age and gender distribution of daily users.

ten-item personality measure (TIPI) [13]. TIPI was developed to assess the big-five personality traits in scenarios where time is limited. It only uses ten items and is therefore suitable to include in an online survey. The survey ended with demographic questions.

3.2 Participants

We recruited participants via Amazon Mechanical Turk (MTurk) resulting in 1,072 completed questionnaires. The MTurk population is diverse in age, education level, and socio-economic status [19]. Thus, a sufficiently large survey has the potential to generalize to a more varied population, compared to traditional recruitment methods with limited geographic diversity [15]. After the removal of outlier responses, according to Tukey's rule [35] and non-smartphone users, 1,003 valid and complete answers remained. Table 1 summarizes the demographics. Participants were mostly US based while 3.5% reported other, varied nationalities. Our sample was representative in terms of gender, with 450 participants identifying as male (44.9%), 549 as female (54.7%), and 4 as non-binary (.4%). Ages ranged from 18 years to 84 years (Mdn = 32). Our sample is closely matched to the average US population in terms of educational levels [5]. Unemployment is above average but 85% of our participants are part of the working population. The distribution of age and gender is shown in Figure 1.

3.3 Results

The findings are listed below. Daily users will be addressed separately.

3.3.1 (Voice) Messaging behavior. We asked participants which messenger applications they used primarily. The most used messenger was Facebook Messenger (62.4%), followed by iMessage (33.8%), and WhatsApp (29.4%). Of our 1,003 participants, 835 (83.3%) reported to have received a voice message before, and 729 (72.7%) reported to have sent a voice message from their smartphone. We divided our participants in five classes of usage frequency: daily users, weekly users, monthly users, yearly users and less-frequent users. The percentages of those groups are shown in Table 2. Usage of voice and text messages differed strongly. While daily users constituted the largest group for text messages, most

	Text messaging		Voice messaging	
	Receive	**Sent**	**Receive**	**Sent**
at least once a day	82.5%	80.2%	20.8%	15.7%
at least once a week	12.0%	12.4%	31.0%	25.3%
at least once a month	3.6%	4.6%	21.4%	18.3%
at least once a year	0.9%	1.3%	5.3%	7.3%
less frequent	1.0%	1.5%	4.1%	5.2%
don't know	0.1%	0.1%	0.6%	0.9%
never used	0.0%	0.0%	16.7%	27.3%

Table 2. Reported usage of participants regarding text and voice messaging. While text messaging usage is highest for daily users, voice messaging usage is highest for weekly users.

participants reported to use voice messages at least once a week. The difference between messages sent and received is also larger in voice messaging than in text messaging.

We analyzed the relationship between the big five personality traits and frequency of sending and receiving voice messages with a Spearman's rank-order correlation. There was a very weak, positive, statistically significant correlation between extroversion and usage frequency (send: $r_s = .127, p < .001$, receive: $r_s = .136, p < .001$). This indicates that extroverted persons are more likely to use voice messages. We also found a very weak, significant correlation between emotional stability and usage frequency (send: $r_s = .100, p < .01$, receive: $r_s = .115, p < .001$). Please note that those r_s values are low and can only explain a small portion of the relationship between variables. Therefore, they need to be interpreted cautiously. We examined age related differences in usage of VMs and thoroughly analyzed different age groups. However, by correlating users age and reported usage frequency we found no salient differences and no indication that voice message usage is more pronounced in specific age groups.

3.3.2 Daily users. We were particularly interested in daily users of voice messaging. When selecting only those from our dataset, 147 users remained. The ranking of most used messengers changes to Facebook messenger (72.1%), WhatsApp (54.4%) and iMessage (24.5%). The huge increase in WhatsApp usage (29.4% to 54.4%) indicates that voice messages are particularly popular on the WhatsApp messenger. Interestingly, while the age distribution stayed similar, the gender distribution changed substantial. Out of the 147 daily users, 90 (61.2%) identified as male and 57 (38.8%) as female, which initially was the biggest group at 54.7%. Not only form male users the majority of daily users, they also report a higher number of voice messages sent per day (male $M = 8.11, Median = 5$, female $M = 5.56, Median = 3$) and received per day (male $M = 7.78, Median = 5$, female $M = 6.91, Median = 5$); and are younger than the female group (male $Median = 30$, female $Median = 37$).

3.4 Motivations for Voice Messaging

We asked participants, who stated to already have recorded a voice message, for what reasons they use voice messages over text messages, leading to 735 responses. Responses ranged from single words (e.g., quicker, easy) to multiple sentences with more than 100 words.

We used thematic coding to analyze the responses, following Robson and Cartan's guidelines [29]. As a first step, two authors familiarized themselves with the collected data. They then used about 50% of the data to generate labels by stacking similar responses. Those labels were then categorized and ordered leading to an initial codebook. The whole dataset was then individually coded by the two authors with this codebook. Remaining uncoded data and unclear

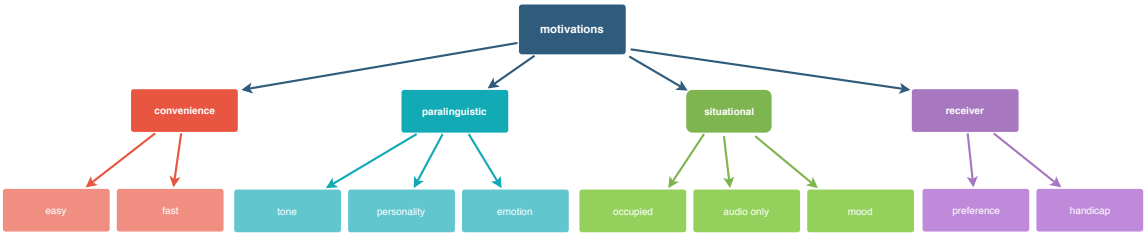

Fig. 2. The simplified codebook. User motivation can be divided into four categories: convenience, paralinguistic, situational restrictions and receiver preferences/handicap.

cases were discussed, leading to a slightly revised and extended codebook. The dataset was then re-coded by both coders. A simplified version of the final codebook is shown in Figure 2. To ensure validity, inter-rater reliability was calculated using a two-way mixed, consistency, average-measures ICC [20] to evaluate the degree of consistency in code occurrences. The calculated ICC was in the excellent range (ICC = .96) [9]. As a final step, remaining conflicts were resolved in a joint session with a third author. 97 responses were excluded due to not being content-related, ambiguous, or the question was misunderstood. During this process of coding we identified the four themes of users' stated rationales for voice messaging that are described in the following.

3.4.1 Convenience. The most mentioned theme (n=359) was the added *convenience* of voice messaging due to ease of use and quickness of recordings. As shown by Ruan et al. [30], voice input via speech recognition is about three times faster than typing and by directly using recordings, voice messages are even faster. Being constrained to a virtual keyboard as is the case with mobile text input is a hurdle for many users. Recurring statements declared that especially long messages are often send via voice messages to avoid having to type a lot of text. One participant told *"Sometimes I don't feel like typing or want to add more detail to a message so I use the voice messaging."* With the amount of things to say, the drawbacks of mobile text input are getting even more severe. Voice messages are used *"when there is a complex problem that need a long, long text message to explain and a solution to said problem."*

Besides those ergonomic features, another interesting aspect that we included in the theme 'convenience' is that users reported being able to express their thoughts more easily when using voice messages. Users reported that *"it's just easier to say what needs to be said rather than taking the time to write it all out"* and it is *"easier to say what you want to say"* via voice messages.

3.4.2 Para-linguistics. The second biggest theme (n=229) that serves as a motivation for voice messaging are *para-linguistic* features that otherwise would get lost in written text. Those features include prosody, pitch, volume, intonation and inflection and are used to make a message more clear. Many users are choosing voice messages because they feel it avoids miscommunication and misunderstandings. A user stated *"it can be too easy for there to be misunderstandings that could result in a negative impact on the relationship"* and that *"much can be lost in translation or through the format"*. By using their voice, many participants feel like it is *"easier to get a point across"*. A participant also stated that *"texts can be misinterpreted and some may think it's something serious when it's not."* Voice messaging is also explicitly used to convey emotions that text cannot transport:

> *"I am old school. It is important to me to hear voice emotions because text isn't very good at conveying emotion. Written things that can convey emotion without voices [are] Novels, Poetry and letter writing but not text [messages]."*

Voice messages are often used for lighthearted messages such as *"relaying a funny anecdote"* or *"say something in a certain silly voice"* but adding a personal touch or making a message more clear was found more frequently in the responses. They are used for particularly important and serious topics where misunderstandings are not an option.

3.4.3 Situational Context. The third category of reasons for voice messages is the *situational context* (n=203) of the user. In many situations, it is more interrupting to type a message compared to just speaking it into the microphone. By just using a single button, it is also a more hands-free option. Many participants stated that they use the voice messaging feature while driving because it is less distracting and allows to *"keep [the] eyes on the road."* While driving is the most often mentioned restricting activity, voice messaging is also used to *"multitask while cooking or doing stuff around the house"*.

One interesting aspect we found is that users stated that they send voice messages in hope to get one back. Some of them explicitly stated that they *"wanted to hear their [friends] voices but not be on a phone call"*. Voice messages are also an effective way of communicating or sharing surrounding and ambient sounds to others. Participants reported to use it for the sounds of their kids and toddlers or just *"something funny that I want to record and send the audio of"*. We also found many cases where participants used voice messaging for things that require audio such as singing happy birthday to a friend or sending music they discovered.

3.4.4 Receiver. Beside the motivations that originated from the sender, we also identified *receiver*-centered motivations (n=34). Those can be divided into preferences and handicaps. Participants reported to use voice messaging for special persons such as family and friends:

> *"My friends and family like to hear my voice [...]. When I voice message I usually am laughing and reacting in a dramatic way which they enjoy hearing."*

Some explicitly mentioned that they use voice messaging because the receiver *"told [them] to use"* it. Voice messages are also used for accessibility and overcoming impairments, for instance if *"someone doesn't like reading, or has bad eyes"*.

The online survey provided us with a demographic overview of voice messaging users, indications that extroverted and emotionally stable persons are more likely to use voice messaging and that daily users are skewed towards younger male users. Furthermore, we discovered and analyzed themes of motivation for voice messaging.

4 VOICE MESSAGING IN-THE-WILD

The online survey took a wide population into account but was limited in scope and relied on self-reporting. To gain more profound insights into the usage of voice messaging, we conducted an in-the-wild study. This allowed us to accurately measure the frequency and times when voice messages are sent. Furthermore, experience sampling is used to enrich the logged data without harming participants privacy more than necessary.

4.1 Apparatus

In order to record all events that are related to voice messaging, we developed an Android application. This application makes use of the fact that the WhatsApp messaging app stores all voice messages in a publicly accessible storage area. When the application we developed is first installed and after every device reboot a background service is started. This service observes the folders where voice messages are being stored by WhatsApp. In case a new message is recorded a new file is created, if a message is played back the file is accessed and if a recording is aborted the file gets deleted.

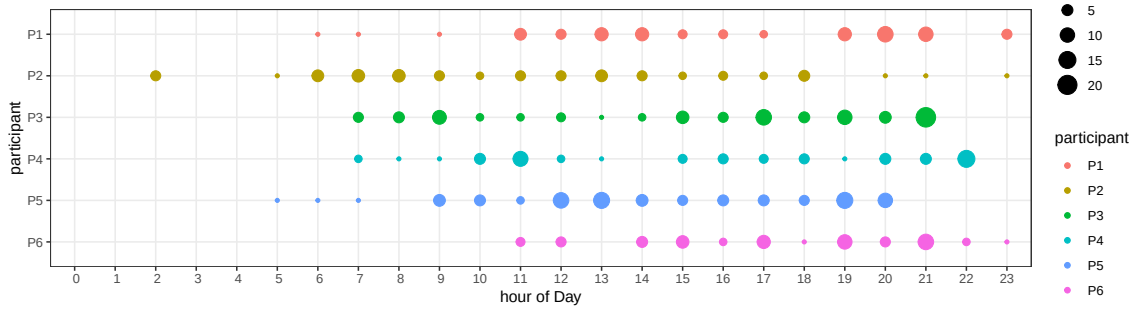

Fig. 3. The event distribution of recording or listening to voice messages over the time of day.

Those file operations trigger events in our service which are then used to create a list of events for each voice file that let us know when it was created, deleted and how often it was played back.

Beside frequency of use, we gathered information about message content, sender/receiver relationship and context of usage. To be able to capture this information in-situ, we incorporated experience sampling into our application. Therefore, on every second sent or playback event, a notification is triggered that presents a questionnaire to the participant asking to answer one or two short questions. We included four different questionnaires that are presented in rotating order to the participants. The first one is asking about the social relationship between sender and receiver. The participant has to select one of the listed group of persons (partner, close friend, family member, acquaintance, room mate, co-worker, boss/superior, teacher/instructor, stranger) or add another option. If the 'other' option is selected, a new name must be specified and this option is added to the list of available options going forward. This option and mechanism was implemented in every questionnaire. Below that question, the 'Inclusion of the other in the self'-scale (IOS) was presented in its continuous form [16]. It shows two circles labeled 'You' and 'X'. Participants need to use a slider to adjust how much these circles overlap, which is a comprehensible tool to assess the closeness of social relationships [1]. The second questionnaire is asking for the location and context of the participants. High-level categories such as 'at home' or 'at work' are given as options as well as an 'other' option, working as described above. A third questionnaire asks for the content of the message. Given options are speech, music, environmental sounds, singing and funny stuff/joking. As a fourth questionnaire the self-assessment manikin (SAM) questionnaire [2] was used to measure emotional responses to voice messages or the mood in which participants have been in when recording a voice message. It uses five distinct figures of the manikin and a 9-point scale for each of the three dimensions: pleasure, arousal and dominance. Beside experience sampling, we used the built-in activity recognition of Google Play Services to label the captured events with the current activity of the user. Therefore, all recognized activities with a probability score of more than 70% were used.

4.2 Procedure

We recruited through word of mouth and snowball sampling 6 participants for the field study. After an initial call, during which the application was installed on a participant's smartphone and the questionnaires were explained, the study lasted two weeks. Once those two weeks passed, participants were reminded to sent a report via a notification within the Android application. After we received this report, we scheduled a closing interview with the participant.

4.3 Participants and Results

To recruit participants for this study authors reached out to persons known to use voice messaging frequently. They were then screened to meet our criteria. Participants needed to be sending VMs at least daily. Additionally they had to use WhatsApp and an Android smartphone because only in such cases the application is able to record the actions of the user. When a participant was found, we used snowball sampling to reach out further. Due to our specific criteria and the privacy relevant topic it was hard to find participants. The six participants which are present resemble the heavy-users of voice messaging. In order to make the results of our study more accessible, we provide an overview of demographics and other personal traits in table 3. During the two weeks, we recorded 846 interaction events within 438 voice messages. The times of day of receipt and recording of voice messages are shown in Figure 3. Thereof, 228 messages were recorded and sent by our participants themselves, 198 messages were received and 12 recordings were aborted. Reasons for abortions were reported as a slip of the tongue (n=5), change of mind (n=4) and by accident (n=2). The duration of messages showed a wide range from 1.46 seconds to 427.26 seconds (7.12 minutes) with a median duration of 17.54 seconds. Users listened to messages they received 1.37 times on average and 0.14 times on average to the messages they recorded themselves. The maximum playback count of a single message was 13 times. When considering the average message duration, the estimated number of words (using an average speech rate of 100 words/minute) is about two times of what was found to be the average message length by Lyddy et al. [18] (29.23 words in voice messages vs 14.3 words in text messages). Even though those numbers are only a rough estimate, they show that in text messaging users try to communicate their content with a low number of words which doesn't help to communicate clearly. In voice messages users are more comfortable to use as many words as needed to communicate their content as shown by the maximum message length of over 7 minutes. Mobile text messages with that much content are very rare.

Interacting with voice messages was perceived as being joyful, as participants rating of pleasure in the SAM questionnaires were in the upper third of the scale for both, sending and receiving (Median = 7.0). The factors arousal and dominance of the SAM questionnaires showed no major tendencies with median values of 4, 4.5 and 5 respectively (see Figure 4). The evaluation of the IOS ratings (0 = no inclusion, 100 = very high inclusion) supports the hypothesis that voice messages are mostly exchanged with close friends and family. The median values are 65 for incoming and 82 for outgoing messages. Both signify a high nearness to individuals with whom voice messages are exchanged. However, it can also be seen that self-chosen receivers were assessed as being closer and participants reported even values of zero nearness for incoming messages (Figure 4).

By using activity recognition we were able to classify each messaging event. Most events occurred when the phone was reported to be 'still' (56.0%) which is rather the general state. In 24.6% of the events the recognition reported 'tilting', which often occurs when a device gets picked up from a desk or a seated user stands up [12]. It signals a change of context such as the end of a specific task. The three remaining recognized activities where 'on foot' (11.0%), 'vehicle' (5.7%) and 'bike' (2.7%).

4.4 Exit Interviews

Following the field study, we conducted semi-structured interviews with our participants. We prepared a script consisting of five parts. After a short introduction, we asked about voice messaging in general, followed by questions about contacts, sentiments, context and two closing questions. All interviews were conducted via Skype and were recorded for post evaluation. Interview sessions lasted 11.43 to 25.00 minutes with a median of 18.13 minutes.

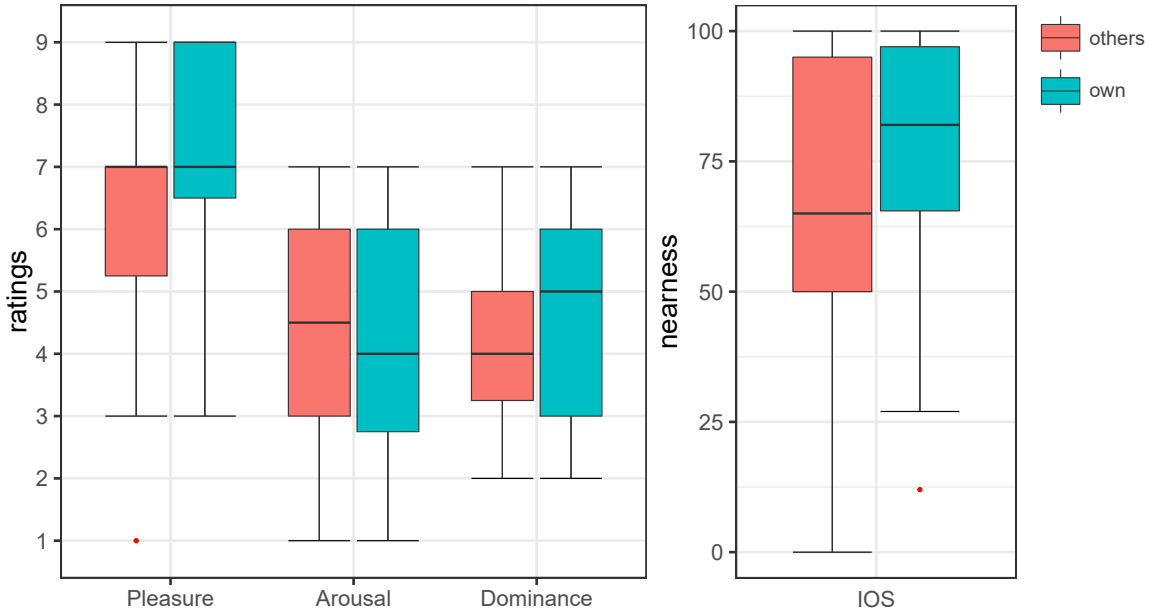

Fig. 4. The scores of SAM questionnaire ranging from 1 (=low) to 9 (=high) and results from IOS-Measure ranging from 0 (=no nearness) to 100 (=very high nearness), both divided into own messages and those from others.

Participant	Age	Gender	Occupation	Relationship	voice messages are:	Send	Received	Aborts
P1	26	male	student	single	convenient	15	55	1
P2	33	male	sales representative	engaged	handy	39	23	2
P3	28	female	physiotherapist	in a relationship	handy	51	41	1
P4	25	male	mechanic	single	-	35	25	3
P5	38	male	radio presenter	married	authentic	59	29	0
P6	20	female	student	single	simplifying	29	25	5

Table 3. An overview to the participants of our in-the-wild study presenting some demographic information and the amount of messages send and received.

4.4.1 General. Half the participants stated that they started to use voice messages around 2014, shortly after voice messages were introduced by WhatsApp. The other three only used voice message for the last one or two years. We further asked them how and why they started to use voice messages. Five of the participants stated that they tried it out and just found it to be convenient, especially when on-the-go, but P5 stated:

> *"I used it because it's a better mean of expression for me. Because with all the emojis, with the smileys and all the stuff, it was just too complicated and too misleading for me. And that's why I used voice messages."*

This quote shows how interwoven usage rationales can be. It includes the great ease of use and para-linguistic features that, while emojis can be misleading [22], help to make a message clearer.

4.4.2 Contacts. When asked how many contacts they regularly voice-message with, all participants reported rather low numbers ranging from 3 to 6 persons. All described those persons as being either very close friends, family members or partners. Three of the participants explained that some of those contacts are distant and voice messaging helps to

stay in touch with them. When asked, how the persons they voice-message with, compare to the overall number of persons they communicate with frequently, all participants stated 20 to 30 persons and one participant stated 50. P2 told about one particular friend:

> *"He has a problem with texts in general anyway, he can't concentrate on reading and that's why he is perfectly suited for voice messages. [...] he usually never writes, he only sends voice messages."*

This shows that voice messaging can be an inclusive means of communication for persons that are not able or willing to text.

4.4.3 Situations and context. In the online survey, secondary occupations where often mentioned as users rationales. When asking about typical use cases for voice messaging, participants replied with the already known activities such as driving, being on-the-go or having full hands. However, those occasions are somewhat user dependent, e.g. not all participants were regular drivers. We further got multiple responses that voice messages are particularly useful for making appointments which is in conflict with the ability to scan texts for important information quickly. As it seems, it is more important to be able to express uncertainty and open choices (usually expressed through a change in intonation such as rising or a fall-rise intonation) than to exchange information in the most efficient way. P3 said:

> *"For example in the evening, when I've planned a lot and I organize the whole thing, for example where to meet and when. [...] then I'm super happy when I can just say something in a voice message: [let's meet] here and there and there"*

We also asked participants how other persons around them influence there voice-messaging usage. While one participant argued that voice messaging has become popular enough to be performed in public, two reported to refrain from usage in public scenarios such as crowded places. Others reported to try to be more discrete but do not refrain from usage.

4.4.4 Formality of voice messages. Voice messages are often perceived as a light-hearted communication form that sits somewhere in between mobile texting and calling. Our participants all stated that voice messages are very informal and relaxed for them. Although they compared it to phone calls, when specifically asked about, almost all of them replied that they wouldn't use them for business communication. Sending a voice message to a superior was considered inappropriate by most participants. Some receive or exchange voice messages with colleagues but only when they are also good friends or get along well with them.

4.4.5 Choosing modalities. By asking participants how they choose whether they use a call, text message or voice message to contact somebody, we found two different approaches. Half of them stated that they decide based on the message and the time needed to get it across. If the message contains just a short information that can be send quickly via text they would use texting. When the message is something more detailed or getting complicated, voice messages are the preferred communication modality. Calls were only taken when timely feedback is crucial, such as emergencies, or if the receiver is unable or refusing to use messenger apps, for instance older relatives. However, the other half of participants stated that the receiver is the most important factor. For contacting strangers, texting is the default but sympathetic or close individuals and family members are almost always contacted via voice messages.

Interestingly, most participants stated that they only rarely call and do not like calling because of the added time and effort. P3 said:

> *"I basically prefer to make phone calls, but it's always like this: Can you reach the person now? Does she really have time?".*

P5 made it more clear by stating:

> *"I think that's just the older generation you're still making phone calls with. With the younger generation you don't call at all anymore."*

and P6 even used calling someone as a tongue-in-cheek threat when friends took to long to answer her messages.

By using our application we were able to investigate 846 events resulting from interacting with voice messages. From this data we were able to gain insight into the practical use of voice messages, e.g. the length of messages and that 5% of recordings were aborted. In situ questionnaires showed that voice messages are perceived as a pleasant interaction and that participants exchange messages with persons who are largely very close to them. Activity recognition determined that almost half of the interactions were in motion or context transitions. In addition, the final interviews allowed us to structure and understand the data collected.

5 LIMITATIONS

Considering the results of our online survey, it is important to keep in mind that MTurkers are a rather tech savvy population and resemble the population of frequent Internet users [19]. This may also be the reason why we did not find salient differences regarding usage behaviour in age groups. Recruiting participants for a two week field study, especially in combination with personal and sensitive data, proved to be difficult. This led to a smaller sample size as desirable and limits the generalizability of the results. It should further be noted that only a small part of the participants in the field study belong to the same population as the MTurk workers.

Although we made sure to implement our application as stable as possible, we cannot guarantee that no events were lost in the process. The Android operating system tries to kill off background services regularly to avoid battery drain and some manufacturers implement their own additional battery saving features. The fragmentation on Android makes it near impossible to test each of these cases. Our service was restarted every time it got killed but messages that are send or received in between may got ignored.

By focusing on users motivation and usage behavior, we left out the users that avoid voice messaging. As important as it is to understand why and how voice messaging is used, as important is it to understand its problems and why some may refrain from usage. This should be further investigated in future work.

6 DISCUSSION AND FUTURE WORK

When we took a closer look at the population of daily users, we found that male users are in the majority and also reported higher amounts of sent messages. While male participants in general were younger, this alone cannot contribute to this imbalance. Previous research [36] found that age and gender is a factor in the acceptance of new technology and especially young men are frequent in the group of early adopters of technology [7]. However, we cannot offer a full explanation and therefore, further investigation is needed.

Analyzing the personality traits of users and correlating them with usage frequency, we found that very active VM users scored higher in extroversion and emotional stability than the rest of our population. Typical characteristics of extroversion are being talkative, active, energetic, enthusiastic and adventurous. Persons that score high on emotional stability tend to be calm, emotionally stable, and free from persistent negative feelings. While the evidence for emotional stability is not as clear, extroversion seems to agree very well with the characteristics of voice messages and their frequent use. Unfortunately, we only used a very brief questionnaire to asses personality features of users but an in-depth investigation of personality traits and correlation with the use of voice features could yield interesting results.

Current media theories often assume that communication is primarily about the exchange of information and availability of dialog partners is not an issue. As an extension of those theories Nardi et al. [24] introduced the concept of interaction and outeraction for modern instant messaging as follows: *Interaction* is the exchange of information as the more technical aspect of communication, *Outeraction* is the social aspect of communication to maintain a sense of connectedness. This categorization can also be used to explain the phenomena of voice messaging. The driving factors in terms of *interaction* are the easy input of voice and the accompanying paralinguistics that preserve emotion and prevent miscommunication. 359 participants stated to use voice messaging out of *convenience*, indicating that by using their voice they could send messages faster, could express themselves more clearly and reported greater overall ease of use when sending a message. Although this is somewhat expected, it is still surprising how strongly this theme is represented. It reminds that text entry on smartphones is far from being easy and worry-free for large parts of the population.

The second most occurring theme, also relating to *interaction*, was participants reporting to prefer voice messages for important and emotional topics, because voice and its *para-linguistic* features were perceived as better preserving meaning, whereas text messages do not properly convey the sender's tone and emotions and therefore are prone to misinterpretation.

In comparison to the main form of mobile communication, texting, voice messaging really excels in *outeraction*. Human speech is able to convey this feeling of being connected for romantic couples or families even when they live temporarily separated. In these situations, simultaneous availability is often challenging e.g. due to differing time-zones. The asynchronity of VMs overcome this problem while still providing intimacy. The SAM ratings recorded during the field study illustrate this intimacy and playfulness of voice messaging as someone speaking out a personal message seems to create a much bigger emotional impact than the receiver just reading a text and seeing emojis attached to it. This can also be used to explain why voice messages are most often used in between individuals that are close to each others and want to keep this closeness. High values on the IOS scale support the participants statements that voice messages are most often exchanged with close friends, family and partners. For exchanging information with others, *outeraction* aspects, that are well supported by VMs, are probably not that important or even undesired. This usage patterns suggests that the individuality and personality of the voice is something that could be further leveraged by allowing users to modify and augment their voice recordings and personalize them similarly to the current use of filters and effects on images (e.g. Instagram). Integrating a (simple) VM editor could provide further processing capabilities and allow for even more expressive storytelling.

While not asked for explicitly, an interesting notion that we discovered in the interviews is that VMs shift the effort necessary for communication towards the receiver, when compared to texting. While texts are hard to enter via mobile phones and their small keyboards, they are rather fast and easy to read for the recipient. For VMs it is he other way round. Composing a message is quick and easy but the retrieval takes more time and effort as reading a text covering the same content. This can lead to interesting usage behaviour in hierarchical structures such as work where it can be consider okay to send VMs from superiors to subordinates but not the other way round. When considering availability and interruptability, we identified problems in VM usage that should be further investigated in future work. In some contexts such as work or social settings participants reported to find it inappropriate or hard to retrieve VMs, due to the increased effort of retrieval. While the content of a text message can be captured by a quick glance, it is not that quick and easy for VMs and therefore, most participants reported to postpone VMs in work settings. VMs are also hard to retrieve when the surrounding soundscape is loud or restricted such as in concerts. We propose to include automated

voice transcripts in mobile instant messengers to make VMs scannable for embedded information such as time and date which should allow retrieval in those constricted situations.

7 CONCLUSION

In this work, we presented insights into the user group of VMs, motivations for and actual usage of smartphone-based asynchronous voice messaging. We first collected a large sample of 1,003 participants to describe the population of voice messaging users, daily users and usage rationals. Using thematic analysis, the four themes *convenience*, *para-linguistic*, *situational* and *receiver* were uncovered as being the major reasons that drive the use of voice messages. To further explore those themes, we conducted a two week long field study. We were able to collect real world user behavior and in-situ questionnaires showed that voice messaging is a joyful interaction and participants exchanged messages with individuals that are, for the most part, very close to them. Activity recognition revealed that nearly half of the interactions are performed when moving or in context transitions and the exit interviews gave us the opportunity to organize and interpret the collected data.

We found that voice messages are a novel communication tool that, when used appropriately and courteously, provides ease of use, helps to handle secondary occupations, and better facilitates maintenance of close relationships.

REFERENCES

[1] Arthur Aron, Elaine N. Aron, and Danny Smollan. 1992. Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology* 63, 4 (1992), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

[2] Margaret M. Bradley and Peter J. Lang. 1994. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry* 25, 1 (March 1994), 49–59. https://doi.org/10.1016/0005-7916(94)90063-9

[3] Tatiana Buhler, Carman Neustaedter, and Serena Hillman. 2013. How and Why Teenagers Use Video Chat. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 759–768. https://doi.org/10.1145/2441776.2441861

[4] Daniel Buschek, Mariam Hassib, and Florian Alt. 2018. Personal Mobile Messaging in Context: Chat Augmentations for Expressiveness and Awareness. *ACM Trans. Comput.-Hum. Interact.* 25, 4 (Aug. 2018), 23:1–23:33. https://doi.org/10.1145/3201404

[5] US Census. 2018. Educational Attainment in the United States: 2018. https://www.census.gov/data/tables/2018/demo/education-attainment/cps-detailed-tables.html Library Catalog: www.census.gov Section: Government.

[6] Barbara L. Chalfonte, Robert S. Fish, and Robert E. Kraut. 1991. Expressive Richness: A Comparison of Speech and Text As Media for Revision. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '91)*. ACM, New York, NY, USA, 21–26. https://doi.org/10.1145/108844.108848

[7] Patrick Y. K. Chau and Kai Lung Hui. 1998. Identifying Early Adopters of New IT Products: A Case of Windows 95. *Inf. Manage.* 33, 5 (May 1998), 225–230. https://doi.org/10.1016/S0378-7206(98)00031-7

[8] Karen Church and Rodrigo de Oliveira. 2013. What's Up with Whatsapp?: Comparing Mobile Instant Messaging Behaviors with Traditional SMS. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 352–361. https://doi.org/10.1145/2493190.2493225

[9] Domenic V. Cicchetti. 1994. Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology. https://pdfs.semanticscholar.org/50d7/f68422d0c0424674f6b235ac23be8300da38.pdf

[10] Henriette Cramer and Maia L. Jacobs. 2015. Couples' Communication Channels: What, When & Why?. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 709–712. https://doi.org/10.1145/2702123.2702356

[11] MAHA El-shinnawy and M. LYNNE Markus. 1997. The poverty of media richness theory: explaining people's choice of electronic mail vs. voice mail. *International Journal of Human-Computer Studies* 46, 4 (April 1997), 443–467. https://doi.org/10.1006/ijhc.1996.0099

[12] Google. 2018. DetectedActivity | Google APIs for Android. https://developers.google.com/android/reference/com/google/android/gms/location/DetectedActivity

[13] Samuel D Gosling, Peter J Rentfrow, and William B Swann. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6 (Dec. 2003), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

[14] Nenagh Kemp. 2010. Texting versus txtng: reading and writing text messages, and links with other linguistic skills. *Writing Systems Research* 2, 1 (Jan. 2010), 53–71. https://doi.org/10.1093/wsr/wsq002

[15] Aniket Kittur, Ed H. Chi, and Bongwon Suh. 2008. Crowdsourcing User Studies with Mechanical Turk. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 453–456. https://doi.org/10.1145/1357054.1357127

[16] Benjamin Le, William B. Moss, and Debra Mashek. 2007. Assessing relationship closeness on-line: Moving from an interval-scaled to continuous measure of including others in the self. *Social Science Computer Review* 25, 3 (2007), 405–409.

[17] Amanda Lenhart, Rich Ling, Scott Campbell, and Kristen Purcell. 2010. *Teens and Mobile Phones: Text Messaging Explodes as Teens Embrace It as the Centerpiece of Their Communication Strategies with Friends*. Pew Internet & American Life Project, Washington, DC. https://eric.ed.gov/?id=ED525059

[18] Fiona Lyddy, Francesca Farina, James Hanney, Lynn Farrell, and Niamh Kelly O'Neill. 2014. An Analysis of Language in University Students' Text Messages. *Journal of Computer-Mediated Communication* 19, 3 (April 2014), 546–561. https://doi.org/10.1111/jcc4.12045

[19] Morgan N. McCredie and Leslie C. Morey. 2018. Who Are the Turkers? A Characterization of MTurk Workers Using the Personality Assessment Inventory. *Assessment* 26 (Feb. 2018), 1073191118760709. https://doi.org/10.1177/1073191118760709

[20] Kenneth O. McGraw and S. P. Wong. 1996. Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1, 1 (1996), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

[21] Denis McQuail. 2010. *Mcquail'S Mass Communication Theory, 5E*. Sage Publications India Pvt Limited, New Delhi. Google-Books-ID: GRtuPwAACAAJ.

[22] Hannah Miller, Jacob Thebault-Spieker, Shuo Chang, Isaac Johnson, Loren Terveen, and Brent Hecht. 2016. "blissfully happy" or "ready to fight": Varying interpretations of emoji. In *Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016*. AAAI press, Palo Alto, 259–268. https://experts.umn.edu/en/publications/blissfully-happy-or-ready-to-fight-varying-interpretations-of-emo

[23] Aarthi Easwara Moorthy and Kim-Phuong L. Vu. 2015. Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space. *International Journal of Human–Computer Interaction* 31, 4 (April 2015), 307–335. https://doi.org/10.1080/10447318.2014.986642

[24] Bonnie A. Nardi, Steve Whittaker, and Erin Bradner. 2000. Interaction and Outeraction: Instant Messaging in Action. In *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work* (Philadelphia, Pennsylvania, USA) *(CSCW '00)*. ACM, New York, NY, USA, 79–88. https://doi.org/10.1145/358916.358975

[25] Midas Nouwens, Carla F. Griggio, and Wendy E. Mackay. 2017. "WhatsApp is for Family; Messenger is for Friends": Communication Places in App Ecosystems. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 727–735. https://doi.org/10.1145/3025453.3025484

[26] Donna J. Reid and Fraser J.M. Reid. 2007. Text or Talk? Social Anxiety, Loneliness, and Divergent Preferences for Cell Phone Use. *CyberPsychology & Behavior* 10, 3 (June 2007), 424–435. https://doi.org/10.1089/cpb.2006.9936

[27] Ruth Rettie. 2007. Texters not talkers: phone aversion among mobile phone users. *PsychNology Journal* 5 (April 2007), 33–57. http://www.psychnology.org/375.php

[28] Felix Richter. 2014. An Average WhatsApp User Sends >1,000 Messages per Month. https://www.statista.com/chart/1938/monthly-whatsapp-usage-per-user/

[29] Colin Robson and Kieran McCartan. 2016. *Real World Research*. John Wiley & Sons, Hoboken, NJ. Google-Books-ID: AdGOCQAAQBAJ.

[30] Sherry Ruan, Jacob O. Wobbrock, Kenny Liou, Andrew Ng, and James Landay. 2018. Comparing Speech and Keyboard Text Entry for Short Messages in Two Languages on Touchscreen Phones. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1, 4 (Jan. 2018), 1–23. https://doi.org/10.1145/3161187 arXiv: 1608.07323.

[31] Lauren E. Scissors and Darren Gergle. 2013. "Back and Forth, Back and Forth": Channel Switching in Romantic Couple Conflict. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, New York, NY, USA, 237–248. https://doi.org/10.1145/2441776.2441804

[32] Lauren E. Sherman, Minas Michikyan, and Patricia M. Greenfield. 2013. The effects of text, audio, video, and in-person communication on bonding between friends. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 7, 2 (July 2013), –. https://cyberpsychology.eu/article/view/4285

[33] Ben Shneiderman. 2000. The Limits of Speech Recognition. *Commun. ACM* 43, 9 (Sept. 2000), 63–65. https://doi.org/10.1145/348941.348990

[34] IBG Tencent. 2017. The 2017 WeChat Data Report. http://blog.wechat.com/2017/11/09/the-2017-wechat-data-report/

[35] John W Tukey. 1977. *Exploratory data analysis*. Addison-Wesley Pub. Co., Reading, Mass. OCLC: 3058187.

[36] Viswanath Venkatesh, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. User Acceptance of Information Technology: Toward a Unified View. *MIS Q.* 27, 3 (Sept. 2003), 425–478. http://dl.acm.org/citation.cfm?id=2017197.2017202

[37] Sarita Yardi and Amy Bruckman. 2011. Social and Technical Challenges in Parenting Teens' Social Media Use. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 3237–3246. https://doi.org/10.1145/1978942.1979422