

VibEmoji: Exploring User-authoring Multi-modal Emoticons in Social Communication

by

Ziqi Zhou

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Mathematics
in
Computer Science

Waterloo, Ontario, Canada, 2022

© Ziqi Zhou 2022

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contribution

This thesis includes first-authored content from the following CHI'2022 publication:

Pengcheng An*, Ziqi Zhou*, Qing Liu*, Yifei Yin, Linghao Du, Da-Yuan Huang, and Jian Zhao. 2022. VibEmoji: Exploring User-authoring Multi-modal Emoticons in Social Communication. *In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 493, 1–17. <https://doi.org/10.1145/3491102.3501940>

The content from this paper has been adapted, revised and extended for this thesis.

Abstract

Emoticons are indispensable in online communications. With users' growing needs for more customized and expressive emoticons, recent messaging applications begin to support (limited) multi-modal emoticons: e.g., enhancing emoticons with animations or vibrotactile feedback. However, little empirical knowledge has been accumulated concerning how people create, share and experience multi-modal emoticons in everyday communication, and how to better support them through design. To solve this, we developed VibEmoji, a user-authoring multi-modal emoticon interface for mobile messaging. Extending existing designs, VibEmoji grants users greater flexibility to combine various emoticons, vibrations, and animations on-the-fly, and offers non-aggressive recommendations based on these components' emotional relevance. Using VibEmoji as a probe, we conducted a four-week field study with 20 participants, to gain new understandings from in-the-wild usage and experience, and extract implications for design. We thereby contribute to both a novel system and various insights for supporting users' creation and communication of multi-modal emoticons.

Acknowledgements

Words cannot express my gratitude to everyone who made this thesis possible.

I would like to express my deepest appreciation to my supervisor Dr.Jian Zhao for his guidance and help. I really enjoy my graduate study time.

I would like to extend my sincere thanks to Dr. Dan Vogel and Dr. Oliver Schneider for being my thesis readers.

I am also grateful to other authors of the paper. It was a wonderful time working with you.

I gratefully acknowledge the help of all the staff of the University of Waterloo.

Table of Contents

Author's Declaration	ii
Statement of Contribution	iii
Abstract	iv
Acknowledgements	v
List of Figures	ix
List of Tables	x
1 Introduction	1
2 Background	4
2.1 Emoticons and Affective Communication	4
2.2 Emerging Multi-modal Emoticons in Current Social Applications	5
2.3 Novel User Interfaces Supporting Communications via Emoticons or Multi-modal Signals	6
3 Scenario	9
3.1 Design Considerations	9
3.2 Usage Scenario and VibEmoji Interface	10

4 VibEmoji System	13
4.1 Preparing Multi-modal Elements: Stickers, Animations, and Vibrotactile Patterns	13
4.2 Gathering Data for Perceived Emotional Properties of Animation and Vibrotactile Libraries	15
4.2.1 Data Collection for Animations	15
4.2.2 Data Collection for Vibrations	16
4.2.3 Post-survey Questions	17
4.3 Recommending VibEmoji Elements	18
4.4 Implementation Details	19
5 Field study	21
5.1 Participants and Study Setup	21
5.2 Data Gathering Methods	22
6 Results and Analysis	25
6.1 Quantitative Results	25
6.1.1 Questionnaire Responses	25
6.1.2 Interaction Logs	26
6.2 Qualitative Results	27
6.2.1 Theme 1: Four Types of Scenarios that Multi-Modal Emoticons Bring Extra Values to Communication	27
6.2.2 Theme 2: Users' Authoring Process—How They Select and Interpret Multi-modal Elements	31
6.2.3 Theme 3: Users' Opinions on the Recommendation Approach	33
6.2.4 Theme 4: Envisaged Opportunities—Multi-modal Emoticons Beyond Mobile Messaging	34
7 Discussion	36
8 Conclusion	40

List of Figures

3.1	(a) VibEmoji enables users to seamlessly create and send multi-modal emoticons during messaging; (b) users could freely combine various stickers, vibrations, and animations to convey rich and nuanced feelings on the fly, in response to the unfolding conversation.	10
4.1	(a) emotional property datasets of stickers, animations and vibrations for providing recommendation based on the elements' emotional relevance; (b) a simplified illustration of the VibEmoji system architecture.	14
4.2	Frequencies of responses on the added values and usage scenarios of sending emojis combined with animations and vibrations.	17
6.1	RS1 to RS6: users' general experience with multi-modal emoticons in comparison with static emoticons; RS7 and RS8: users' general style in authoring multi-modal emoticons (7-point Likert scale, 1=strongly disagree, 7=strongly agree).	25

List of Tables

2.1 An overview of emerging multi-modal or user-customizing features of emoticons in current social applications (customization refers to whether users could modify the properties of emoticons, rather than importing new sets of stickers)	5
5.1 Participant information	22
6.1 Four types of scenarios that multi-modal emoticons suggested extra values to communication experience. Each type is explained by an exemplar case gathered from the participants, with the multi-modal emoticons illustrated. More examples could be found in Section 6.2.1.	28

Chapter 1

Introduction

The form of social communications has dramatically evolved since the advance of technology. Online social communication is ubiquitous in daily life, and, the use of emoticons plays an essential role in online communications. Emoticons are the universal and cross-cultural visual language and overcome linguistic and cultural barriers in communication. Additionally, they are a way for people to express themselves concretely by expressing sentiments, thoughts, or nonverbal cues that are difficult to put into words [10]. The use of emojis is constantly increasing over the last decade on Twitter [6] and around half of the comments contain emoji in Instagram [13].

Along with the fast-growing popularity of emoticons, new emoticons still continue being demanded by users for various communication needs [17]. Meanwhile, the classic, static emoticons seem to no longer fully suffice people's various needs for expression. Users continue seeking further enriched, more expressive, and customized ways to communicate via emoticons. The term "multi-model emotions" was used to represent different physical inputs and outputs modalities, including video, speech, and sound. The "multi-model emoticons" in this research will focus on the modalities of the emoticon itself, including movement, haptics, and visual effects. The trend for more expressive, flexible and improved forms of emoticons is inspired by the emergence of multi-modal emoticons and adopted by major online social softwares:

- Apple allows users to create emojis with customized appearances (Memoji [2]) and record animated emojis (Animoji [48]).
- Google Gboard's Emoji Kitchen [18] allows users to combine two emojis to create a new one.

- Tencent WeChat [46] and Huawei MeeTime [24] display certain emojis with animation and haptic effects.
- Telegram [45] allows users to interact with a sent emoji by adding animation and haptic effects.

Though many industry products have started to explore the use of multi-modal emoticons, the related research in the HCI field are insufficient. There is little research evidence and knowledge about people’s creation process, usage and perception of multi-modal emoticons in everyday communications. Thus the thesis aims to address this untapped potential.

The thesis will introduce VibEmoji, a user-authoring multi-modal emoticon system that supports users to effortlessly author, and communicate via multi-modal emoticons during everyday messaging. In order to provide new empirical information about the use and user experiences of multi-modal emoticons, we utilize VibEmoji as both a design to study about and a technology probe to study with, generating new empirical knowledge about the usage and user experiences of multi-modal emoticons, and extract relevant implications for future research and design in this domain. Notably, VibEmoji has several features that go beyond or set it apart from previous design examples:

1. The current consumer products, for example, Telegram and Apple iMessage, fix the combination of emoticon stickers, animation effects, and vibrotactile feedback, and users have no way to modify or regenerate the combinations of these elements to express their meanings. On the other hand, users have complete freedom to choose from and mix three diverse sets of elements: pictograph emoticons, animation effects, and vibrotactile patterns in VibEmoji.
2. The current systems always provide automatic selections for users; for example, the smile emoji will be suggested when a user inputs “happy,” which might make irrelevant choices since the user’s purpose was misunderstood. On the other hand, in VibEmoji utilizes a more passive suggestion technique without restricting users’ options. In VibEmoji, when a user picks the first element, for example, an emoticon, the system then predicts the remaining two sets of elements (vibrations and animations) and updates the display order to give certain components precedence on the interface. The prediction algorithm uses the data from two sources: (1) frequency of user-created combination patterns, and (2) emotional significance between the selected components using the valence-arousal model [40].

3. Multi-modal emoticons in the current applications must be pre-defined, not at the time of usage. For example, users have to set the configurable parameters for Apple Memoji or record animations for Apple Animoji before using them. In contrast, VibEmoji allows user-authoring multi-modal emoticons in live communication. This feature enriches the user experience and lets the researcher explore the multi-modal emoticon creation process based on users' feelings and context.

The study uses the emotional properties of emoticons gathered by [39], but there are no sufficient datasets for vibrations and animations. In order to solve this, two questionnaire surveys were conducted to collect the perceived emotional qualities of 15 animation effects (52 respondents) and 60 vibrotactile patterns (52 respondents).

VibEmoji contains a mobile application with the corresponding backend system and was used in a four-week field study test with 20 participants divided into 10 pairs. Each participant pair consists of friends or partners who have previously engaged in regular smartphone texting. In order to reveal empirical knowledge about how the participants developed, utilized, and experienced multi-modal emoticons in their daily conversations, a mixed-method approach was employed to collect both quantitative and qualitative data. As a result, VibEmoji makes two main contributions:

1. A system that facilitates user creation and usage of multi-modal emoticons in online communication;
2. Empirical knowledge about how people actually create, share, and experience multi-modal emoticons in their everyday lives, along with relevant implications for future designers and researchers.

Chapter 2

Background

2.1 Emoticons and Affective Communication

While the most famous early example of emoticons, `:-)`, has been proposed by Fahman [16] in his email to colleagues in 1982, the earliest use of emoticons in computing systems is dated back to the 1970s (the PLATO system [5]). While the term *emoticon* and *emoji* are often used interchangeably (as this thesis also does), the name emoji is originally rooted in the Japanese mobile market (firstly supported in a product by J-Phone in 1997), which literally translates as “facial letters/characters.” As a universal, powerful pictographic language, emoticons/emojis have been widely used to concretely and conveniently convey emotional feelings that would otherwise take more words to articulate.

For this reason, a great number of related studies analyzed the usage of emoticons among relatively large-scale samples to understand people’s emotions in various contexts: e.g., to detect software developers’ emotions in communication [9], to collect students’ emotional states in learning [54], or to understand people’s political attitudes [21], to name but a few.

Besides analyzing affective patterns that can be commonly extracted from people’s emoticon usage, another body of research focuses on the contextual and personal aspects, to uncover how people assign new meanings to emoticons: e.g., repurposing or appropriating emojis beyond their originally intended meanings [52, 27], expanding nonverbal expressions to reduce the dependency of texts [58], or developing highly customized usage [20].

Studies encompassing emoticons and emotional communication often used theories and models of emotion as their basis for measures and analyses. Overall, two types of theories

Table 2.1: An overview of emerging multi-modal or user-customizing features of emoticons in current social applications (customization refers to whether users could modify the properties of emoticons, rather than importing new sets of stickers).

Application	Animation	Haptics	Recommendation	Customization	Customize On-the-fly
Facebook Messenger	–	–	– ¹	–	n/a
Apple Memoji/Animoji	++	–	+ ¹	++ ³	–
Apple iMessage Effects	++	++	–	–	n/a
Google Gboard	+	–	+ ¹	+ ⁴	+
Telegram Messenger	++	+	–	–	n/a
Tencent WeChat	++	+ ²	+ ¹	–	n/a
Huawei MeeTime	++	+ ²	–	–	n/a

Notes:

++ : fully implemented;

+ : partially implemented;

– : not implemented

¹ Emoji recommendation based on current text input.

² repetitive and simple vibration.

³ pre-configured avatars and pre-recorded animations.

⁴ feature mash-up of two Emojis.

were most frequently referred to: discrete emotion theories, which generalize emotions into discrete categories [11], and dimensional emotion theories, which model emotions in continuous dimensions: e.g., valence-arousal model proposed by Russel and Barrett [40]. For instance, Rodrigues et al. established the Lisbon Emoji and Emoticon Dataset (LEED) [39], in which they collected 505 participants’ perception of 238 emoticons, including the dimensions of valence (positive—negative), and arousal (arousing—calm). These emotional properties of emoticons could inform both the analysis of emoji usage and the development of emotional interfaces. LEED was also used in this study in developing a recommendation algorithm for multi-modal emoticons.

2.2 Emerging Multi-modal Emoticons in Current Social Applications

While emoticons have gained enormous popularity in social communications, the traditional, static emoticons seem no longer to suffice people’s various needs for self-expression. Users continue seeking more customized, further enriched forms of emoticons. Many social applications have developed their unique sets of emoticons (e.g., Facebook Messenger [15], Telegram Messenger [45]), and a few of them have also explored certain user-customization features (e.g., Apple Memoji [2], Google Gboard [18]). More importantly, current social applications start to support (yet limited) multi-modal features in emoticons (see Table 2.1). These emerging features of popular social applications suggest that emoticons are evolving

from simply static visual pictograms into more dynamic compounds of multi-modal elements. Meanwhile, we could expect such multi-modal emoticons to afford more customization space for users to create richer and more engaging ways of expressing themselves and empathizing with others.

However, as [Table 2.1](#) summarizes, current social applications have not fully leveraged the potentials of user-customization in multi-modal emoticons, leaving several meaningful opportunities for new design explorations. First, while animation and haptic feedback are increasingly incorporated, current systems only enable pre-designed, fixed combinations of animation and vibration, either applied on specific emoticons (e.g., WeChat [46], MeeTime [24]), or provided as rigid options (e.g., iMessage Effects). None of the systems allow users to freely combine different animations, vibrotactile patterns, and emoticons to create new meanings. Second, current systems (e.g., Memoji [2], Gboard [18], WeChat) only explored the recommendation of emojis based on users' text input. None of the existing cases has explored how to recommend multi-modal combinations for users. Third, few systems have been designed to facilitate users' customization on the fly. Memoji and Animoji, for example, both require users to pre-define the avatar or pre-record animations to prepare customized stickers, instead of spontaneously creating and sending new multi-modal emoticons according to the unfolding conversation. Although Google Gboard supports the mash-up of two Emojis during a conversation, it does not enable on-the-fly modification of animations or haptic feedback.

The above opportunities have underlain the core design features of the VibEmoji system, which (1) supports users to freely combine various emoticons, vibrotactile patterns, and animations as multi-modal emoticons, (2) provides appropriate recommendations about relevant multi-modal elements, and (3) enables users' on-the-fly customization during a conversation. By embodying these design features that extend beyond existing cases, and evaluating the design in the wild, we aim to extract new empirical knowledge that could advance the development and user experience of multi-modal emoticon systems.

2.3 Novel User Interfaces Supporting Communications via Emoticons or Multi-modal Signals

To the best of our knowledge, prior work has rarely explored supporting users to author and communicate via multi-modal emoticons. However, rich novel interfaces have been created and studied in the HCI community, to facilitate users' communications via (visual) emoticons or multi-modal signals (e.g., haptic experience). These design cases have served

as inspirations for our exploration.

A stream of research has been focused on exploring new ways for users to communicate via (pictorial) emoticons or emojis. Opico [28] enables emoji-first communication: users could respond to each other using sequences of emojis, expressing feelings or simple concepts without texts. MojiBoard [1] is a keyboard that eases users' entry of parametric emojis: a series emojis that can convey emphasis or micro-stories. A number of cases targeted fully automating users' process of selecting or sending emojis. ReactionBot [33] captures users' facial expressions and accordingly adds emojis to their text messages on Slack. Face2Emoji detects users' facial emotional expressions for quick emoji selection [12]. Other explorations concerned automating the selection of emojis/emoticons based on emotion keywords [49], sentences [30], or speech signals [23]. Voicemoji [55] explored voice-based emoji entry for visually impaired users. A few studies tackled challenges for the use of emojis by improving emoji accessibility for visually impaired users [47] and reducing the cultural barriers to understanding emojis [31]. Increased attention has been paid to user customization of emoticons [19], for example, in generating new emojis based on users' sketch and text input [34], or allowing two intimate users to co-customize their emoticon shortcuts (DearBoard) [20].

A few studies suggested the promises of integrating animated representations to enlarge the communication capacity of (static) emoticons. For instance, Animated GIFs were found to afford rich interpretation and nuances in nonverbal communication [26]. However, in GIFs/short videos, the image and motion effects are fixed combinations, whereas in this study, we aim to enable users to freely combine static emoticons with different motion effects. As related, the AniSAM study suggested that adding animated representation to static icons could more effectively visualize emotional states (e.g., arousal) [44]. This suggests adding animated effects might enrich the nuances and expressiveness of static emoticons. Harrison et al. [22] proposed KineticIcons, a rich library of kinetic behaviors that can be applied on static GUI elements such as icons, to enable extra communication affordance. Our study utilizes Kinneticons to establish a versatile set of animation components for multi-modal emoticon authoring.

Touch, or haptic experience, is an essential type of non-verbal cue for social communications. Besides the abundant haptic studies on supporting users to understand, monitor and operate technologies (e.g., [14, 7, 53]), prior work has extensively explored technology-mediated haptic experience in interpersonal communications. For instance, vibrotactile feedback was leveraged in both mobile and wearable devices, for affective communication in distance, by resembling the feeling of touch, cheek-poke, or handshaking between users: e.g., [4, 57, 37, 36, 51]. A recent case explored supporting mediated touch in face-to-face settings without breaking social distancing rules during COVID [56]. Several studies have

also encompassed supporting haptic communication in text messaging, including a communication API to support the vibrotactile feedback in mobile messaging [25]. In addition to end-user interfaces, related research has contributed to theories [29], tools [8, 43], or evaluation techniques [41] of haptic experience to better support related design and research. For example, VibViz [43] offers a large, diverse library of vibrotactile stimuli for haptic feedback design, based on which we developed the vibrotactile library for VibEmoji. In summary, the prior studies illustrated the values of haptic modality in enabling affective communication and social connectedness [35, 51, 42]; and a few specifically suggested benefits of combining haptic signals with visual cues [57]. These conclusions supported our decision of leveraging haptic feedback as one of the major components for multi-modal emoticons.

Chapter 3

Scenario

In this chapter, we present the design considerations, a usage scenario, and detailed features of the VibEmoji interface.

3.1 Design Considerations

The design rationales behind VibEmoji are underlain by our exploratory research objectives: to extract knowledge about how people create, share and experience multi-modal emoticons in daily communication, and how to better design user-authoring multi-modal emoticon systems. Therefore, as addressed in [chapter 2](#), the core features of VibEmoji are designed to extend beyond, or differentiate from existing design cases, to better facilitate the generation of new understandings in this domain. Here we briefly consolidate our design considerations for these core features, before going into the details of the interface design:

D1: Enabling users to freely combine multi-modal elements. Existing systems that support multi-modal emoticons only provide a limited number of fixed vibration-animation effects for users to customize their emoticons. To probe how to support deeper customization and richer expressions by users, we design VibEmoji as an open-ended authoring interface that allows for freely combining various emoticons, vibrotactile patterns, and animation effects.

D2: Providing recommendation without limiting users' options. Current emoticon systems have explored automating users' selection by recommending an emoticon-based on their textual inputs (keywords). However, to enable users' nuanced, differing



Figure 3.1: (a) VibEmoji enables users to seamlessly create and send multi-modal emoticons during messaging; (b) users could freely combine various stickers, vibrations, and animations to convey rich and nuanced feelings on the fly, in response to the unfolding conversation.

expressions, VibEmoji explores an alternative recommendation approach that facilitates, rather than fully automated, users’ choice. Instead of recommending the best solution, the interface prioritizes the most relevant multi-modal options based on the users’ present selected element, while still keeping all other options available. This supports both the user’s efficient authoring of classic combinations and open exploration of new expressions.

D3: Supporting on-the-fly authoring of multi-modal emoticons. The user-customization process of existing emoticon systems is often carried out separately from the actual use during online chatting. Users often need to pre-configure emoticon features or dynamic effects in order to use them later. VibEmoji is designed to support users’ on-the-spot authoring of multi-modal emoticons during online communication. It streamlines users’ authoring process into four steps: selecting three multi-modal elements and press the “send” button (see Figure 3.1 (b)). With VibEmoji, we are able to probe how users may create ad-hoc, improvised expressions based on the unfolding conversation.

3.2 Usage Scenario and VibEmoji Interface

Here we present an example of how users could use the VibEmoji mobile application to combine emoticons, vibrotactile patterns, and animation effects to create multi-modal emoticons on the fly, during communication. Anna and Brad are chatting online. Brad said something really funny which has made Anna burst into laughter. Anna wants to fully

express how much she enjoyed it, so she decides to send a “laughing tears” emoji 😂 and enhance its expression with animation and haptic effects. She opens the VibEmoji mobile application first, then presses the VibEmoji icon next to the text input area. Thus, she unfolds the VibEmoji keyboard, and selects that emoji, as well as a vibrotactile pattern that feels like body shaking, and animation of bouncing up and down. Anna is satisfied with this combination, so she presses the “send” button, and this emoticon enhanced by the animation effect and vibrotactile pattern is sent to Brad’s device. Brad receives it and knows that Anna finds what he said funny. He wants to let Anna know that he feels the same and is still laughing about it. So he pressed that emoticon from Anna, and its haptic feedback is rendered again with the animation on both of their devices. Anna feels this response from Brad.

As [Figure 3.1](#) (a) shows, the VibEmoji interface looks similar to an Emoji keyboard, and can be easily unfolded during messaging by pressing a small icon next to the text-input field. But different from a conventional Emoji keyboard which only shows emoticons, the VibEmoji keyboard has three segments to display three types of multi-modal elements: emoticons, vibrotactile patterns, and animation effects (**D1**). As [Figure 3.1](#) (b) shows, each segment could be scrolled horizontally to browse its all elements. Users could make selections from the three sets of elements regardless of the selection order (**D3**), e.g., they could start from an emoticon like Anna did in the scenario, but they could also start with selecting a vibrotactile pattern or animation effect first. Users could press an element to select it, and it will be highlighted on the interface ([Figure 3.1](#) (b)). When pressed again, that element will be deselected.

In the animation segment, all animation effects loop continually so that users could easily preview them. When no emoticon element is selected, the animation previews will be applied on a default neutral icon; otherwise the animation previews will be rendered using the currently selected emoticon. In the haptic pattern segment, each vibrotactile element is displayed by a thumbnail preview that visualizes its waveform (based on its intensity and sharpness parameters) in a simplistic manner. These waveform previews are designed to help users quickly see the general trend and length of a vibrotactile pattern. Each time when selected, a vibrotactile pattern will be rendered by the haptic module of the mobile phone. If there is a selected animation, the vibration will be rendered in synchrony with the animation. This helps users directly preview what their current combination feels like and eases their quick experimentation with different options.

To further assist users’ seamless creation on the spot, VibEmoji uses an unobtrusive recommendation technique to predict and prioritize the options that are more relevant to the user’s current selection (**D2**). Namely, each time when the user makes the first selection, be it an emoticon, vibrotactile pattern, or animation effect, the display order of

the elements from the other two segments will be updated accordingly. The elements that are predicted as more likely to be combined with the current selection, will be put forward on the display, so that they can be more easily found.

This recommendation follows two-fold principles. First, elements that were frequently combined with the current selection and sent by the user in the past, will be prioritized by the display. On top of that, elements that have more similar perceived emotional properties to the selection (based on their scores in the valence and arousal dimensions [40]), will have higher priority on the display. For example, assume that the user’s first selection is an emoticon scored relatively high in both valence and arousal, i.e., perceived to convey both “positive” and “exciting” emotions. Then, the elements in the vibrotactile and animation segments which convey similarly positive and exciting feelings will be given higher priority on the display, to better match the present communication intention of the user. This way, VibEmoji offers a non-aggressive recommendation approach: on the one hand, it eases users’ selection by prioritizing options that are more often used or have more emotional relevance to the preceding selection; on the other hand, it still guarantees users abundant space to explore and experiment with new multi-modal combinations. The input data and algorithm of the recommendation technique will be detailed in the next chapter.

Chapter 4

VibEmoji System

In this chapter, we report how VibEmoji is developed to support the aforementioned design features, including the preparation of the multi-modal elements and data collection of their emotional qualities, as well as the scheme and implementation of the system.

4.1 Preparing Multi-modal Elements: Stickers, Animations, and Vibrotactile Patterns

Each VibEmoji emoticon can be authored by a user by combining three multi-modal elements: a (static) sticker, an animation effect, and a vibrotactile pattern (**D1**). These stickers, animation, and vibration sets are intended to be open-ended, meaning that users could add new elements to expand each set (e.g., a user could include new stickers to the sticker set). We have prepared default elements for the three sets (50 stickers, 15 animations, and 60 vibrations), which affords a great number of combinations possible for creating multi-modal emoticons (also see diverse examples gathered from the field deployment discussed in [Section 6.2.1](#)).

Stickers: The default set consists of 50 Apple emojis (Version IOS 10.0). We chose Apple emojis as our default stickers due to their frequent usage among mobile users. The chosen 50 stickers (see [Figure 3.1](#)) were selected based on an emoji usage survey by the Unicode Consortium¹, which classified all the Unicode emojis based on usage frequency. We filtered them out and chose the frequently used facial emoji stickers (e.g., not objects,

¹<https://home.unicode.org/emoji/emoji-frequency/>

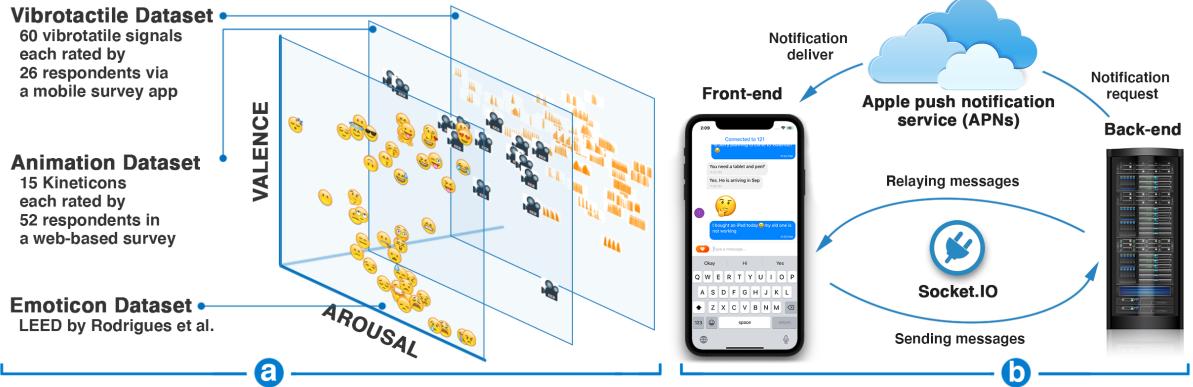


Figure 4.1: (a) emotional property datasets of stickers, animations and vibrations for providing recommendation based on the elements’ emotional relevance; (b) a simplified illustration of the VibEmoji system architecture.

stars, etc.). This set of emoticons has also been targeted in Rodrigues et al.’s survey [39], due to their usefulness and frequent usage for expressing a variety of emotions.

Animations: The default set consists of 15 animation effects from the work of Kinecticons [22]. It provides a set of diverse, open-ended and multi-purpose kinetic effects to be combined with different graphical elements (e.g., icons) to support a wide range of communication purposes [22]. Hence, these animations are not strongly bonded to certain semantic meanings or connotations. This open-endedness is in accordance with our design, which is to combine the animations with different stickers to afford rich expressions. Of the original 39 Kinecticons, we excluded the animation effects concerning two objects, which are not applicable in our case. We also filtered out those that could not be appropriately applied on circular emoji stickers, e.g., because they were designed for rectangular icons or menus.

Vibrations: The 60 default vibrotactile patterns were chosen from the work of VibViz [43], a diverse vibration library intended for the design to convey information through the haptic channel in digital devices. VibViz was chosen also because of its potential possibilities to be combined with different stickers and animations to convey various meanings. Of the original 120 vibrations, we excluded the ones that were longer than 10 seconds, because they are not suitable for short, emoticon-based instant-messaging scenarios. We further filtered out vibrations that had less emotional relevance to the default set of stickers in terms of valence and arousal (further explained in Section 4.2.2).

4.2 Gathering Data for Perceived Emotional Properties of Animation and Vibrotactile Libraries

As mentioned, our design rationale is to grant great freedom for user authoring while keeping the authoring process intuitive and effortless (**D1** and **D3**). To achieve this, we use a recommendation algorithm that predicts most likely combinations based on the selected element, and accordingly updates the display order of the un-selected sets, to ease the choices of users. Such prediction is based on two kinds of data: the perceived emotional properties of each element (i.e., valence and arousal [40]), and the user’s historical usage data. As the prerequisite of the recommendation algorithm, the datasets of perceived emotional properties need to be established beforehand. For the sticker set, the perceived emotional properties of the core collection of Apple emojis have been already measured by Rodrigues et al.[39] (with a sample size of 505), and this measured collection has covered the frequently used emojis we chose based on the Unicode survey. We therefore utilized their open-sourced data for the algorithm. However, for the animations and vibrations, adequate datasets remain to be established. We thereby conducted questionnaire surveys to gather the data.

4.2.1 Data Collection for Animations

A web-based questionnaire survey was conducted with 52 respondents to gather the perceived emotional properties of the 15 Kineticon animations. Kineticons were designed with a rich vocabulary of kinetic behaviors to convey diverse intentions and meanings [22]. However, no prior work has assessed what emotional feelings these animations could trigger from users, which is crucial for emotion-based recommendation.

The questionnaire was specifically developed for data collection using `Javascript` and `Node.js`, which renders each animation effect on a separate page and asks respondents to use 7-point Likert scales to rate the valence and arousal of each animation (Supplementary Material A). The animations are rendered using the neutral default icon used in the original user study of Kineticons [22]. The order of the animations was randomized for each respondent.

The respondents were recruited from the Amazon Mechanical Turk (MTurk) platform (aged from 25 to 54; 27 males, 24 females, and 1 other). Each survey session took 10–20 minutes from a respondent. The compensation for each respondent is \$2.5. 1 invalid response were excluded due to incompleteness. In the end, 52 responses were used (each of

the 15 animations was rated by 52 respondents) as the prerequisite data of the animation set (see [Figure 4.1](#)).

4.2.2 Data Collection for Vibrations

The original VibViz study results consist of the valence and arousal scores of each vibrotactile pattern rated by three researchers [43]. However, no survey has been conducted so far, to collect more data about users' perceived emotional properties of the VibViz patterns, which had become part of our development tasks. Besides excluding long vibrations ($> 10s$), we filtered out vibrations that were relatively further to an emoticon sticker from the sticker set in the valence-arousal space. These more distant vibrations had less emotional relevance to the chosen emoticon stickers, therefore, they are deemed less likely to be combined together with the chosen stickers.

For each sticker in our sticker set, we marked the 5 closest vibrations in the valence-arousal space, and accumulated the mark frequency for each vibration. In the end, there were 60 variations marked more than once, which resulted in our final vibration element set.

A second questionnaire survey was conducted with 52 respondents for the emotional perception of the vibrotactile patterns. Due to the fact that respondents need to access a device with a haptic engine to experience the vibrotactile patterns, we developed an ad-hoc survey application on the iOS platform using Swift UI (Supplementary Material B). To avoid each survey session getting too long and causing fatigue or boredom in respondents, we split the 60 vibrotactile patterns into two surveys (30 for each). Each survey took 15–20 minutes for each respondent. In the survey application, to rate each vibrotactile pattern, the respondent was asked to first press a button to experience the vibration before its Likert scales appeared. This button was available on-screen throughout the rating of that pattern, so that the respondent could replay the vibration anytime when needed.

Due to the consideration of device access, respondents were only recruited through word of mouth, which were required to have access to an iPhone 8 or above (to make sure the haptic engine works smoothly). The survey app was distributed to the respondents through both local build and TestFlight². In the end, 52 qualified respondents (aged from 18 to 59; 35 males and 17 females) completed the survey and each vibrotactile pattern was rated by 26 respondents (see [Figure 4.1](#)).

²<https://developer.apple.com/testflight/>

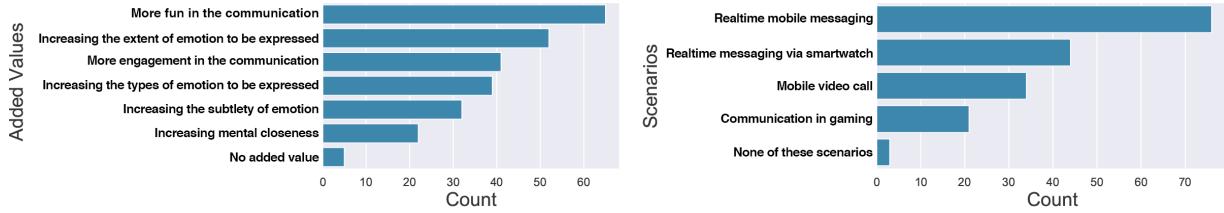


Figure 4.2: Frequencies of responses on the added values and usage scenarios of sending emojis combined with animations and vibrations.

4.2.3 Post-survey Questions

After rating the emotional properties collected for the recommendation algorithm ([Figure 4.1](#)), the 104 respondents of the animation and vibration surveys filled in two extra multiple-choice questions: if you could send emojis combined with animations and vibrations in online communication, (1) what would be the added values that you foresee, and (2) what would be the meaningful usage scenarios for it? The choices of these questions were formulated based on our literature study ([Section 2.3](#)) and pilot interviews with five college students. Open questions were not used to avoid the extra questions demanding too much time and effort from the respondents.

The results are shown in [Figure 4.2](#). For added values, the most frequent response is more fun in the communication (65 times). These responses inspired the questionnaire developed for our field study (see [Section 5.2](#)), to see whether similar opinions would emerge from lived experiences with multi-modal emoticons. In terms of added values, the most frequent response is more fun in the communication (65 times). The second frequent is the increased extent of emotion to be expressed (52), followed by more engagement (42), accuracy of emotion expression (41), types of emotion to be expressed (39), and subtlety of emotion (32). Increased mental closeness was much less chosen (22), except for no added value (5). These responses inspired the questionnaire developed for field study users (reported below), to see whether similar opinions would emerge from lived-experiences with multi-modal emoticons. In terms of meaningful scenarios, most frequent option is real-time messaging via mobile phone (76 times), followed by real-time messaging via smartwatch (44), mobile video call (34), communication in gaming (21), and none of these scenarios (3). This confirmed that mobile messaging is a meaningful scenario for exploring multi-modal emoticons.

4.3 Recommending VibEmoji Elements

As demonstrated in [Section 3.2](#), to ease the authoring process of multi-modal emoticons, VibEmoji offers an unobtrusive recommendation to predict and prioritize the multi-modal emoticon elements (i.e., stickers, animations, and vibrotactile patterns) that are more relevant to the user’s current selection (**D2**). Initially, before any user selection, all the three elements have default display orders on the VibEmoji interface. This mimics the convention of most emoji keyboards where the location of each emoji is fixed to facilitate users with consistent access based on their memory. Once a selection is made by the user, the undecided elements are ranked based on their association to the selected element by the user. Given an undecided element u in one modality M_u , and a selected element s in another modality M_s , the recommendation algorithm considers the following two essential factors to obtain a ranking score for u .

First, elements that have more similar perceived emotional properties to the selected element are prioritized. This is because it is less appropriate to combine elements that are quite different in their emotional properties, which may result in expressing contradictory emotions in a single VibEmoji emoticon. Based on the gathered data described in [Section 4.2](#), we simply use the Euclidean distances between elements in the valance-arousal 2D plane, like those in previous studies [43, 39]. Specifically, the perceived emotional similarity P is defined as:

$$P(u, s) = \frac{1}{\sqrt{(u_v - s_v)^2 + (u_a - s_a)^2}}, \quad (4.1)$$

where $u_{v,a}$ and $s_{v,a}$ represent the valence and arousal values of u and s , there is no case for $u_v = s_v, u_a = s_a$ in gathered data.

Second, elements that were frequently combined with the current selection and used in the past are prioritized. This is based on the assumption of recently visited items will be likely revisited again, which is widely used in a range of recommendation systems such as Google Search, Amazon, etc. However, simply counting the frequency of u and s used in combinations overlooks the overall usage patterns. We thus employ a TF-IDF approach [38] that is commonly used in information retrieval. The TF (term frequency) reflects the frequency of u and s used together with respect to u being used with all the elements in the modality M_s . That is,

$$TF(u, s) = \begin{cases} \frac{F(u,s)}{\sum_{s' \in M_s} F(u,s')} \\ 0, \text{ if } \sum_{s' \in M_s} F(u,s') = 0 \end{cases}, \quad (4.2)$$

where $F(u, s)$ is the number of occurrences that u and s are combined in VibEmoji emoticons by the user in the past. Further, the IDF (inverse document frequency) reduces

the bias that the element u is used too often with all the elements in the other modality M_s , which promotes the diversity of the recommendation. A document is a point in the valance-arousal 2D plane. Thus,

$$IDF(u, s) = \log \frac{n}{\sum_{s' \in M_s} N(u, s')} + 1, N(u, s) = \begin{cases} 0, & \text{if } F(u, s) < 1 \\ 1, & \text{if } F(u, s) \geq 1 \end{cases}, \quad (4.3)$$

where n is the number of elements in M_s . Therefore, the final TF-IDF score is $TF-IDF(u, s) = TF(u, s) \cdot IDF(u, s)$.

We then use a weighted score to compute the ranking score for u with respect to s by combining the two factors:

$$R(u, s) = \alpha P(u, s) + \beta TF-IDF(u, s), \quad (4.4)$$

where α and β are weights. In our development, we set $\alpha = 0.6$ and $\beta = 0.4$ based on our empirical observation. We compute this ranking score for every element $u \in M_u$ and sort the scores descendingly to obtain the display order of the elements on the VibEmoji interface. When there are only one selected element (e.g., a sticker), we reorder the elements in the other two modalities (e.g., vibration and animation) based on the above ranking method. When there are two selected elements (e.g., a sticker and a vibration), we simply use the average of the two ranking scores to obtain the final score of each element in the third modality (e.g., animation).

4.4 Implementation Details

In this section, we introduce how we built the VibEmoji system in detail. As shown in [Figure 4.1](#), the system contains a front-end mobile application to support on-the-fly authoring of multi-modal emoticons, and a back-end server to perform data processing, messaging, and communication.

The front-end application was implemented based on the React Native framework, because it can be easily adapted to other platforms in the future (e.g., Android), in addition the Apple iOS platform this project has been built upon. To have fine-grained control over how the vibrotactile patterns are played on the device, we developed a custom React Native bridge to expose the native iOS haptic API. Our implementation enabled the app to play and pause haptic patterns defined in AHAP (Apple Haptic and Audio Pattern) files, which were identical to the ones used in surveys. For the animations, we recreated the selected Kinections (see [Section 4.1](#)) using `react-native-reanimated`. We used

`react-native-gifted-chat` to implement the front-end messaging application, which provides a comparable look and feel to other messaging applications (see [Figure 3.1](#)).

The back-end server was developed using `Node.js`, which is responsible for relaying messages between users and delivering notifications. To increase the robustness of the system, the server stores a queue of undelivered messages when the receiving party is offline, and re-sends these messages once the users are re-connected. Specifically, notifications on iOS are managed using APNs (Apple Push Notification service). We used `socket.IO` for sending and receiving messages between the front-end and the back-end for reliable real-time communication. To enable sending and receiving VibEmoji emoticons, we created a custom encoding in plain text behind the rendered multi-modal emoticons. For each emoticon, the front-end sends the encoded string to the back-end which relays the message to the receiving party. Then, the front-end on the receiving side re-renders the message based on the encoded string.

Chapter 5

Field study

Using VibEmoji as both a design to study about, and a technology probe to study with, we conducted a four-week field evaluation to gather empirical knowledge about how people create, share and experience multi-modal emoticons in daily online communications, and thereby surface relevant design implications for future research and development.

5.1 Participants and Study Setup

Twenty participants (in 10 pairs) were recruited for a four-week field evaluation. The participants were aged 19–36 ($MD=29$, $IQR = 10.5$), with 12 females and 8 males. They are residents from three countries: Canada (10 participants), China (8 participants), and the Netherlands (2 participants) (see [Table 5.1](#)). Each pair of participants had already established a social relationship with each other; namely, they had known each other for 10 months to 22 years ($MD=4$ years, $IQR = 2$). Participants are referred to as P# in the following text, and consecutive numbers indicate the pairs (e.g., P1 & P2; P3 & P4). Among the 10 pairs, nine of them were friends, and one pair was significant others (P15 and P16). More details about the participants could be found in Supplementary Material C. Participants were recruited via the word of mouth based on two basic requirements: (a) they already use an iPhone 8 or above as the primary mobile phone so that the built-in haptic engine ensures the same sensorial quality of the designed vibrotactile elements; and (b) they can be paired with another person that they already know and have regular communications via online applications.

In the beginning of the study, each participant installed the VibEmoji mobile application on their mobile phone through Apple TestFlight. To ensure that they could experience

Table 5.1: Participant information

Participant	Age	Gender	Country of Residence	Pair No.	Pair Relationship	Length of Relationship
P1	30	F	the Netherlands	Pair 1	Friends	5 years
P2	31	F	China	Pair 1	Friends	5 years
P3	30	F	China	Pair 2	Friends	7 years
P4	29	F	China	Pair 2	Friends	7 years
P5	32	F	Canada	Pair 3	Friends	4 years
P6	32	F	Canada	Pair 3	Friends	4 years
P7	36	F	China	Pair 4	Friends	1.5 years
P8	34	F	China	Pair 4	Friends	1.5 years
P9	35	M	China	Pair 5	Friends	4.5 years
P10	30	M	the Netherlands	Pair 5	Friends	4.5 years
P11	21	M	Canada	Pair 6	Friends	3 years
P12)	21	M	Canada	Pair 6	Friends	3 years
P13	19	F	Canada	Pair 7	Friends	3 years
P14	20	M	Canada	Pair 7	Friends	3 years
P15	21	F	China	Pair 8	Significant others	10 months
P16	29	M	China	Pair 8	Significant others	10 months
P17	22	M	Canada	Pair 9	Friends	4 years
P18	21	M	Canada	Pair 9	Friends	4 years
P19	24	F	Canada	Pair 10	Friends	22 years
P20	23	F	Canada	Pair 10	Friends	22 years

the system in a naturalistic manner, they were asked to use the application in their daily lives to communicate with each other as they already did prior to the study. No structured sessions of chatting were required by the researchers. Neither were the participants asked to use the system exclusively: they were encouraged to use VibEmoji, while they could still use their existing online communication tools at the same time. It was reminded that some risky information might need to be avoided when using the application to chat, such as bank accounts or passwords. Some common and safe topics were provided as examples: including updates about each other's recent life, recent news from the public media, or planning future activities together. Meanwhile, it was made clear to the participants that these were just examples, and they could discuss any topics they desired. Each participant was compensated \$15 after their participation.

5.2 Data Gathering Methods

Given our highly explorative goal, we adopted a mix-method approach to gather both quantitative and qualitative data in various forms during the field evaluation period. Our data gathering methods included unstructured feedback from the participants throughout the four weeks, interface usage logs generated on the back-end of the system, a questionnaire completed by each participant, and a 30-minute in-depth interview with each pair, which

are detailed below.

Unstructured feedback. During their whole period of participation, the participants were encouraged to share their experience and thoughts to the researchers anytime as desired, through multiple types of communication channels, such as voice calls, voice messages, or textual messages. Along with such feedback, they also used screenshots or screen recordings to share some usage examples of multi-modal emoticons that they found meaningful or interesting. These quick and unstructured feedback sessions were transcribed and annotated for analysis.

System logs. Participants' usage logs stored in the back-end were partially used for generally understanding how they interacted with the interface. These retrieved system logs included the number of messages or VibEmoji's they had sent, as well as their operations on the VibEmoji interface: e.g., when and what they pressed on the interface while selecting the multi-modal elements to construct a VibEmoji. To be noticed, the textual contents of participants' chat history were not directly retrieved by researchers for analysis, unless the participants proactively shared a certain part of their chats in unstructured feedback or in final interviews.

Questionnaire. Towards the end of their participation, each participant was provided with an online questionnaire via Google Forms, asking about their general experiences of using VibEmoji. In the questionnaire, we asked about their general styles of constructing multi-modal emoticons: e.g., in which order they made selections across the three types of multi-modal elements, and whether they had discovered several classic combinations and reused them in chat. Moreover, we asked about their general experience of VibEmoji in comparison with the conventional emoticons, in terms of conversation engagement, fun, expressiveness, etc. (see [Figure 6.1](#)). These data were meant to gather their high-level experience and general opinions which could serve as supplementary or triangulation to the in-depth interview data.

Semi-structured interview. A semi-structured interview was conducted for each pair, to gather vivid, in-depth empirical data about how they used VibEmoji to create and share multi-modal emoticons, and what their detailed experiences were. Each interview took 30–45 minutes and was structured in three sections.

- *Section 1 (8 minutes):* The pair was asked to describe their general experience of using VibEmoji. They were also asked to give more explanation about the general experiences they rated in the final questionnaire.
- *Section 2 (15 minutes):* The pair was asked to provide concrete examples about the multi-modal emoticon they created and considered meaningful in the specific context.

In addition, they were asked to reflect on their detailed workflows, such as how they selected different elements to construct new multi-modal emoticons, and how they experience the various design features.

- *Section 3 (7 minutes):* To probe their latent needs about multi-modal emoticon systems, the pair were asked to envision what important features the next version of VibEmoji should have. To probe future design opportunities, the pair was asked to envisage what future scenarios can benefit from multi-modal emoticons, in addition to instant messaging.

In the end, an open discussion is encouraged for participants to share additional thoughts. All interviews were audio-recorded and transcribed verbatim for a thematic analysis along with the unstructured participant feedback data.

Chapter 6

Results and Analysis

In the four-week field study, we collect the system interaction logs, unstructured feedback, questionnaire, and interview results. The quantitative results from the questionnaire and interaction logs and qualitative results from unstructured feedback and interview are presented.

6.1 Quantitative Results

6.1.1 Questionnaire Responses

RS1 to RS6 in [Figure 6.1](#) show the 20 participants' overall experiences with the VibEmoji emoticons, in comparison with the static emoticons they had been using before (on a 7-point Likert scale). In general, they considered that the multi-modal emoticons had contributed to the engagement (RS1), fun (RS2), and expressivity (RS3) of the conversation, and helped

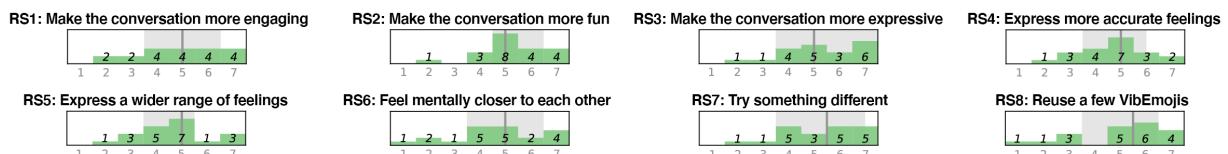


Figure 6.1: RS1 to RS6: users' general experience with multi-modal emoticons in comparison with static emoticons; RS7 and RS8: users' general style in authoring multi-modal emoticons (7-point Likert scale, 1=strongly disagree, 7=strongly agree).

them to express more accurately (RS4), and a wider range of feelings (RS5), and increased their mental closeness (RS6) to each other: with all medians residing on 5 ([Figure 6.1](#)). As shown by the detailed rating distribution in the figure, it seems that users gave higher ratings (i.e., 6 or 7) towards multi-modal emoticons' contributions to the engagement, fun, expressivity, and mental closeness, in comparison with their contributions to the accuracy or the range of feelings.

Moreover, VibEmoji affords great freedom for users to combine various multi-modal components. To generally probe whether such freedom of exploration was meaningful, the questionnaire asked about their general style of usage: i.e., RS7 and RS8 in [Figure 6.1](#). As the figure shows, overall, the participants tended to try something different (variations of combinations) when authoring multi-modal emoticons in chatting (RS-7: MD=5.5). Meanwhile, they also tended to reuse a few combinations (RS-8: MD=5.5). This suggests that while being explorative in use, the participants had also formulated a few frequently used expressions (see examples in [Section 6.2.1](#)), which verifies our design choice for supporting both frequency-based recommendation and open exploration (**D2**).

The open-ended design of VibEmoji allows users to flexibly decide the order of selecting multi-modal elements. When asked about participants' habituated order, 19 preferred stickers to be the first element to choose, while only one preferred starting from vibration. Divergence emerged among the users, in terms of the second element they preferred to select. While 13 participants would go for animations, six would select a vibration, and the other would go for stickers. More analysis of the users' selection order and decision-making in the authoring process is discussed in [Section 6.2.2](#).

6.1.2 Interaction Logs

During the whole period of the field evaluation, the participants used VibEmoji in as daily communication software, they sent 1,824 textual messages and 581 multi-modal emoticons using VibEmoji: approximately every three text messages were accompanied with one VibEmoji. Per pair of participants, the median number of textual messages sent is 86 (IQR = 162, MIN=21, MAX=954), and for multi-modal emoticons, the median is 47.5 (IQR = 49, MIN=15, MAX=182). The logs also provided a rough insight into the time spent for the users to author a multi-modal emoticon. We defined an interaction timeframe that started from the users' first operation on the authoring interface to the moment of pressing the “send” button. The median of the interaction timeframes is 7.09 seconds (IQR = 13.01). A slight difference could be seen, when comparing the timeframes of the first 10% (MD=9.6 seconds, IQR = 18.1) VibEmoji emoticons sent with the later 90%

(MD=7.1 seconds, IQR = 12.7) across all pairs. This may indicate a learning stage in the beginning, and the users became faster after being familiarized with the interface. Moreover, given that users had been actively exploring different variations throughout the study (RS7 in [Figure 6.1](#)), these timeframes should also reflect their time spent for exploration and experimentation. There is no strong pattern regarding the distribution of usage of multi-modal emoticons. This indicates the multi-modal emoticons could be seamlessly integrated into textual communication.

6.2 Qualitative Results

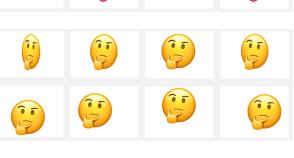
A thematic analysis was conducted to analyze the qualitative data gathered from the semi-structured interviews with the participants, as well as their unstructured feedback provided to us throughout the field study. This inductive analysis method was chosen due to our purpose of establishing a set of structured, systematic meanings (themes) [3] about how the participants created, shared, and experienced multi-modal emoticons in the wild with the VibEmoji system, in order to generate rich empirical knowledge for future research and design. Our analysis followed the six-phase procedure detailed by Braun and Clarke [3], which included familiarization with data by organizing notes and annotations across the dataset, generating initial codes based on research objectives, searching for themes by establishing connections among codes, reviewing and finalizing themes, and re-contextualizing the themes to formulate findings. In this section, we address these qualitative findings, in light of some quantitative results reported in the prior section.

6.2.1 Theme 1: Four Types of Scenarios that Multi-Modal Emoticons Bring Extra Values to Communication

This theme presents vivid usage examples to illustrate what multi-modal emoticons were created and shared in the wild and how they meaningfully benefited the participants' daily communication experiences. We thereby contextually examine the added values of multi-modal emoticons in addition to conventional (static) emoticons, to offer new empirical understandings about multi-modal emoticons.

In general, the participant agreed that using multi-modal emoticons in their daily communications had contributed to their online chatting experience. For instance, as P16 experienced, *“overall, it is pretty nice, it’s a good extension of [a current social application]. In real-life, human emotions are multilayered, and our facial expressions are very*

Table 6.1: Four types of scenarios that multi-modal emoticons suggested extra values to communication experience. Each type is explained by an exemplar case gathered from the participants, with the multi-modal emoticons illustrated. More examples could be found in Section 6.2.1.

Scenario	Exemplar Case	Vibration	Emoticon and Animation Effect
Scenario Type (i): Enhancing Existing Meanings of Static Emoticons	P4: using a “snore” vibration to enhance 😴		
Scenario Type (ii): Creating New Meanings beyond Original Emoticons	P16: appropriating 😱 with a “struggling” effect to express “I don’t want to shut up”.		
Scenario Type (iii): Setting Atmospheres for Unfolding Conversation	P15: the multi-modal combination often used to set an atmosphere for making jokes.		
Scenario Type (iv): Constructing Attentive, Empathetic Responses	P8: the same emoticon with varying motions and vibrations conveyed attentive, non-repeating responses, without interrupting the conversation partner.	 	

dynamic [...] These different options [vibrations and animations] could deliver different emotions. It’s richer [than conventional emoticons]. It adds more dimensions for communicating emotions.” Similarly, as P5 felt, “*It [VibEmoji] is convenient, easy to use. With vibrations and animations added, the emoji feels more three-dimensional. You could feel more emotions from it.*” It has also been suggested that both heavy users and less-frequent users of existing emoticons, could perceive values of using multi-modal emoticons: e.g., “*I send a lot of stickers every day, this [VibEmoji] could expand my expressions of emotions*” (P3); and “*I don’t use a lot of emojis when chatting, but this type of emojis [multi-modal emoticons] could express richer feelings. Because the emoji has animation and vibration, it has more [information] in it.*” (P19). Correspondingly, the participants reported abundant usage cases, and identified different types of meaningful uses of multi-modal emoticons.

Scenario Type (i): Enhancing Existing Meanings of Static Emoticons. The participants provided examples that they used animation effects and vibrotactile patterns to enhance or animate the existing meaning of an emoticon, and made it more vivid or expressive in the conversation.

For instance, as Table 6.1 shows, P3 once asked P4 if she went to sleep already. P4 said not yet, but she would sleep soon, and then she sent a multi-modal emoticon with the *sleepy face* Emotji 😴. To enhance this Emotji, she used an animation and a vibration that both could be associated with the behavior of sleep: the animation is spinning slowly

and slightly, and the vibration “*felt like snoring.*” Both P3 and P4 thought that this combination was expressive and made the conversation more fun. The next example is from P15: when chatting with P16, she augmented the Emoji *face blowing a kiss* 😘 with vibrations that felt like the sounds of kissing but had different duration. She further explained that she would use a shorter vibration to represent a light kiss, and a longer vibration for a long kiss. As another example, P7 sent P8 a VibEmoji based on the emoji *smiling face with heart-eyes* 😍, to express the excitement of seeing a celebrity. To enhance the expression, she used the animation of rapid shrinking and expanding, and a vibration felt like heartbeats.

Scenario Type (ii): Creating New Meanings beyond Original Emoticons. In these examples, the participants flexibly combined the multi-modal elements to convey new meanings that were additional to, or different from the original meaning of the used emoticons. And oftentimes, such new messages or expressions cannot be simply conveyed by the base emoticon alone.

One example was reported by P12, in which he asked P11 to play an online game but did not get a response. He then sent a multi-modal emoticon combining an animation of rapidly waving from side to side, with the Emoji *grinning squinting face* 😜, and a vibration that matched the animation. He selected this “*vibrant*” motion in order to “*urge him [P11] to reply*”, whereas the Emoji was not selected for specific reasons and just to afford a positive connotation. Another creative case was shared by P16 (Table 6.1). He appropriated the meaning of the Emoji *zipper-mouth face* 😮 by combining it with a “*naughty*” and rapid motion, and an intense vibration, which both felt like ”struggling, to express that he did not want to stop talking. As he explained, “*This emoji [zipper-mouth face] originally means ‘shut up’, but [with the selected animation,] now it has become ‘I don’t want to shut up’. It then had extra meaning.*” More such examples were offered by the participants. For instance, P17 combined the Emoji *hugging face* 😊 with a motion of bouncing up and down, to convey a “*creepy*” impression when joking with P18. P2 combined the Emoji *grimacing face* 😬 with an animation and a vibration that felt like “*trembling with cold*” when discussing the weather with P1.

Above examples demonstrate how users’ could creatively construct new expressions and assign new meanings in daily communications when they can freely combine multi-modal emoticons, which confirmed our D1.

Scenario Type (iii): Setting Atmospheres for Unfolding Conversation. Participants found they could construct multi-modal emoticons to (re-)set the atmosphere for a conversation session. In these examples, the prior sent multi-modal emoticons which left on the chat display continued contributing to the ambience of the communication, and

reminding conversation partners of certain vibes during a chat session.

A concrete example was from P15. She had a particular multi-modal emoticon for starting a joke, which included the emoji *winking face with tongue* 😜, an animation of zooming in and out, and an intense vibration (Table 6.1). As she described, “*Every time I want to make fun of [P16], or joke with him, I will send this. It’s like an opening, so that he won’t get mad at me. It sets the tone for the chat [...] a reminder that I’m joking.*” P9 also mentioned that he used the multi-modal emoticons to create relaxing vibes: “*I sent VibEmoji [emoticons] because I don’t want the conversation to be formal or distant. I hope to have some humor, or closeness in the atmosphere.*” Such continuous rendering of a certain atmosphere was partially attributed to the animation components of the multi-modal emoticons, which were continuously visible to users as long as the emoticon is displayed. As P7 reflected, “*the animation itself created a sort of atmosphere [...] It’s playful [...] hard to describe, but it’s contagious from person to person. We both [P7 and P8] could get it.*”

Scenario Type (iv): Constructing Attentive, Empathetic Responses. regards constructing multi-modal emoticons to formulate attentive, empathetic responses. In these examples, the participants created multi-modal emoticons as spontaneous responses to their partners in the unfolding conversation. As experienced by their conversation partners, these responses constructed on the spot could convey more attentive and empathetic feelings than responses via simply standard or preset emoticons.

P8, for example, was once listening to P7 talking about her work. During listening, she also responded to P7 by sending two multi-modal emoticons that both included the emoji *thinking face* 🧐, but had different animations and vibrations which all felt relatively slight and calm (Table 6.1). As she explained, compared with using a static emoticon, or sending the same multi-modal emoticon twice, the spontaneous variations she created on the fly could better convey the meaning “*I am listening with care*”, since “*it shows that I have received your messages and I give you different response each time, rather than simply copying [the same responses].*” And this meaning was indeed conveyed to P7: “[P8] sent me the stickers of thinking, with slowly flipping or spinning [...] felt like she was still thinking and thinking about different things I said [...] She was expressing ‘I am listening’ and she did not want her expression to be interrupting [...] compared with static stickers, this response is more interactive, instead of repeating. It creates richer nuances and feels more friendly.”

Another example was from P3 and P4 when P4 said that her father was going to have a surgery, P3 asked more about her father’s situation and responded to her with a multi-modal emoticon built upon the crying face emoji 😢. Then P4 also sent a

multi-modal emoticon using the same Emoji. What made them feel empathized with each other was that they also used a very similar vibration to express sadness: “*It felt nice that we used the same vibration [...] we shared the same feeling [...] we both thought this vibration could express sadness.*” Such empathetic experience was also reported by other pairs. For example, P1 and P2 appreciated the “*tacit understanding*” (P2) or the “*the sense of empathy*” (P1) between them, when they used similar animation or vibration element to express similar feelings in the conversation: “*we could associate with same emotions [...] don't know why but we just can [...] it's not fully conscious.*” In similar cases, P5 also experienced “*resonance*” with P6. As put by P8, when responding to each other using multi-modal emoticons with certain similar components, “*it feels like we are in sync [...] a sense of being connected.*” Above examples suggest the great values of multi-modal emoticons in enriching users’ spontaneous non-verbal responses to each other, and enhancing their sense of connectedness and interactiveness in the conversation.

As shown in the examples, these benefits could only be afforded when users are able to conveniently and flexibly construct multi-modal emoticons on the fly, during the unfolding conversation, which contextually confirmed **D3**.

6.2.2 Theme 2: Users’ Authoring Process—How They Select and Interpret Multi-modal Elements

This theme elaborates on participants’ process of creating multi-modal emoticons. Namely, we summarize patterns from participants’ rich explanations concerning their preferred orders and strategies of making selections, as well as their (common and differentiated) interpretations of the multi-modal elements. To the best of our knowledge, no prior knowledge has been gathered regarding how users value and prioritize the different types of components when constructing a multi-modal emoticon. To probe a better understanding of this, we designed the VibEmoji interface to be open-ended: users could make selections across the three types of components in any order.

Which Element to Select First (and Why): According to the users’ reports (from both the interview and questionnaire data), a recurrent pattern of their selection order can be seen: in many cases, users would select an emoticon first, and then an animation or a vibration (although meaningful exceptions were also reported). Detailed accounts have been provided by them. First, as agreed by the participants, emoticons often served as the base to set the tone for their multi-modal expression. As P5 stated, “*emoticons decide the theme of the emotion to express*”, or, as P18 put, “*emojis are the major carrier of what I want to express.*” In contrast, animations and vibrations were often used as nuanced

modifications or further enhancement to the base meaning afforded by the emoticons. As they reasoned, this could be because emoticons are more direct and explicit, while animations and vibrations are more open-to-interpretation: e.g., “*when texting, I already associated my feeling with some emoji*” (P2) but “*when selecting from vibrations, I do not have a particular one in my mind*” (P20). Compared with animations, vibrations are even less concrete, since they are “*more subtle*” (P19) and “*not so visual*” (P20).

However, according to the participants, sometimes, animations could have the first-place contribution to the expressed meaning. For instance, P9 gave an example that he first decided to use the animation *moving left and right* which felt like “*nudging him [P10] to wake him up*” to ask for P10’s response. Then he chose the emoticon 😎 only because it was fun. A few similar examples were also told by other participants. In addition, P3 reported a different pattern: oftentimes, her second step is selecting a vibration, before looking for an animation. An extra criterion mentioned by several participants is that they sometimes tended to select animations and vibrations that could better match each other’s temporal properties: such as rhythm or periodic duration. e.g., P15: “*I would make the vibration match the motion. [P16 added:] it’s like dubbing a movie.*”

Vibrations were Associated with Everyday Sounds, while Animations were Associated with Body Language: Another noticeable pattern that emerged from the users’ descriptions is how they used associations to explain their decision-making on selecting vibrations and animations. Namely, when reasoning about why they chose certain vibrations, they often associated their choices with certain everyday sounds; whereas, when explaining their choices for animations, they often used associations with body language. For instance, P3 explained that when constructing a multi-modal emoticon to convey greetings to P4, she selected a vibration that felt like the sound of “*knocking on the door*.” P14 used a vibration to express a strong sad feeling because it felt like an “*electrical noise*.” As another example, P15 explained that she chose a vibration that felt like the “*tick-tock*” sound of a clock, to match the slow circular animation she selected. P9 also mentioned an example in which his selected vibration reminded him of the funny background music both he and P10 were familiar with. Other examples also included associations between vibrations and the sounds of snores, heartbeats, or human voices.

Besides these, rich examples were gathered about how the animations were interpreted as body language. For example, P19 explained that she combined an animation of swinging left and right with a not-so-happy face 😒 to express disagreement because it felt like “*shaking head left and right, disagreeing*.” P17 added an animation of slight and quick shaking to the emoji *grinning face with sweat* 😅 to express the meaning of embarrassment, since the motion resembled the “*face twitching when someone is embarrassed*.” As both P17 and P18 agreed, the animations could make the expressions “*specific and*

concrete [...] more like adding body movements.” P9 similarly confirmed that when looking for animations, he did take “*body languages*” into considerations, either consciously or subconsciously.

Diverse Individual Interpretations: In addition to the common pattern of associating vibrations and animations respectively with sounds and bodily movements, individual differences also widely existed in the participants’ interpretations. For instance, while P4 mentioned that she preferred to choose longer, more intense vibrations which felt richer and more interesting to her, P20 thought that “*long vibrations feel angry.*” For another example, while P12 considered that bigger motions could be used to represent happier feelings, P19 was concerned that bigger motions sometimes could appear too exaggerating and less honest. As P5 pointed out, such differences: “*could have something to do with [a user’s] personality, as well as the topic of the conversation.*” These individual differences in users’ decision-making again, confirmed that it is meaningful to ensure a certain amount of open-endedness in design and allow users to freely combine various multi-modal emoticons based on their personal and contextual interpretations (**D1**).

6.2.3 Theme 3: Users’ Opinions on the Recommendation Approach

This theme summarizes users’ experiences and comments on the recommendation approach of the VibEmoji system, including the two-fold recommendation principles and the non-aggressive recommendation strategy.

Why the Two-fold Principles were Meaningful: The VibEmoji recommendation algorithm follows two-fold principles. One principle is to prioritize the multi-modal options that were more emotionally relevant (in terms of valence and arousal levels) to the user’s selected element. The participants considered this principle to be intuitive and oftentimes meaningful. As confirmed by P15, “*it’s pretty accurate, and it’s similar to what I thought [...] for active and happy emojis, I tend to add faster and bigger motions.*” P8 also thought that “*stronger emotions match stronger vibrations, weaker emotions match weaker vibrations.*” They considered this principle to be helpful especially when users did not want to spend much time browsing the options. As stated by P19: “*some users are lazy, like me. I would use the recommendation. The options on the first screen are already sufficient*”

The other principle is to prioritize the elements that have been more frequently combined with the current section by the user in the past. The participants perceived this to be practical and efficient. For instance, to P17, “*historical frequency is pretty helpful, it’s quite often that I reused some combinations.*” And P3 felt the frequency-based principle

“pragmatic” and “smart”: “*it’s personalized, so it’s about my own feelings.*” In our design, the initial recommendations are based on emotional relevance, since no user data are gathered. Yet, over time, the frequency will shape the interface more and more, so that it would be adaptive to each user’s personal preferences. This design decision has been explicitly confirmed by the participants. As P5 pinpointed, the emotional relevance could provide meaningful guidance to the users especially when “*in the beginning, no pattern [of use] has been formed*”, and after a certain time, it is reasonable to “*meet each user’s different needs, and be more personalized.*” A very similar comment was also given by P16.

Advantages of Non-aggressive Recommendation: Following **D2**, the VibEmoji interface features a non-aggressive recommendation strategy that prioritizes the potentially relevant options, without automating users’ decisions. In the interviews, the participants commented on this feature in comparison with some relatively more “assertive” recommendations they experienced from similar systems. For example, an emoji keyboard that predicts the “best” result to automate users’ selection based on their text input. Overall, the participants had recognized two major advantages of the non-aggressive recommendation strategy over the more assertive ones in terms of multi-modal emoticons. First, the non-aggressive recommendation ensures users’ personal expressions. As P18 put, “*if everyone is using this [recommended option] then why do I have to use it? I want to express my own emotions [...] I won’t send the same one each time. I could try something new.*” By prioritizing relevant options, users could be facilitated while still preserving sufficient variations for creating their own expressions. Second, an algorithm might not be successful all the time, due to its limited access to the whole conversation context. As P18 argued, a recommendation “*might not always be precise in predicting what you really want to express.*” As telling examples for this, in [Section 6.2.1](#), the participants reported a series of cases about how they combined multi-modal elements creatively to express meanings that are quite subtle, or different from the original meaning of the selected emojis. In this sense, a non-aggressive recommendation strategy could avoid giving arbitrary or irrelevant suggestions to users, while leaving space for users to make the modifications. These experiences thereby contextually verify our **D2**.

6.2.4 Theme 4: Envisaged Opportunities—Multi-modal Emoticons Beyond Mobile Messaging

This theme summarizes participants’ envisions about desirable future scenarios or new application domains that multi-modal emoticons (emoticons augmented by animation effects and vibrotactile patterns) could be meaningfully leveraged, in addition to the mobile context.

Interestingly, a number of their envisaged applications went beyond human-human communication. For instance, P11 envisaged that in an AI-enhanced smart home system, a voice agent (e.g., Cortana by Microsoft, or Alexa by Amazon) could also utilize multi-modal emoticons to communicate with human users: “*apart from the voice interactions, if it could send me emojis with animation and vibrations [e.g., on mobile devices], it would feel different, it would feel more human, and richer information could be communicated..*” P3 also envisaged a smart home context, in which the Internet-of-Things products could more expressively interact with users via multi-modal emoticons, and she thought a smart-watch could be a meaningful medium to receive vibrotactile signals. P4 from the same pair talked about a public setting: “*In some shopping malls, there are now AI robots for searching foods and places. If these robots could display emoticons with animations and vibrations, they would be more attractive to younger users I guess.*”

Moreover, P15 suggested using a multi-modal emoticon to summarize people’s attitudes from internet reviews, e.g., of a restaurant: “*instead of reading all the reviews, you could then first feel the general reaction of people who went to eat, from an emoji with animation and vibration.*” P7 and P16 similarly envisaged the usage of multi-modal emoticons in combination with the “barrage captions” in live-streaming, or live videos (a type of dynamic captions that are created by viewers in real-time and moving across the video to enable momentary communications between the live-streamer and viewers or among viewers). As P16 states, “*current live barrages are mainly texts, if dynamic emoticons and vibrations are added, it might be more fun and interactive.*” In addition, both P5 and P12 envisaged that future authoring tools of multi-modal emoticons might consider supporting users to tap the touchscreen to define new vibrotactile patterns, or drawing a trajectory to create a new animation to further unleash users’ creativity.

Chapter 7

Discussion

This paper focuses on multi-modal emoticons: a trend burgeoning in recent social applications which enhances traditional static emoticons with animations or vibrotactile feedback. Multi-modal emoticons are increasingly introduced in current social applications. However, little empirical knowledge has been accumulated concerning how people create, share and experience multi-modal emoticons in daily communications. Addressing this opportunity, we have designed and evaluated the VibEmoji system to gather empirical understandings from real-world usage and experience, and surface design implications for supporting users' creativity and communication of multi-modal emoticons.

As shown in our field deployment, VibEmoji enabled a user-driven exploration. In various scenarios, the participants creatively combined stickers, animations, and vibrations to author multi-modal emoticons on the fly, which helped their spontaneous affective communication in the unfolding conversation. As indicated by the results, using multi-modal emoticons could make online chatting more fun, engaging, and expressive, increase their mental closeness, and potentially help them to communicate more specific, or richer emotional feelings. These experiences have been further supported and concertized by the participants' detailed descriptions in the qualitative findings. Beyond presenting a novel design case and contextually confirming the benefits of multi-modal emoticons, the rich empirical data also enabled us to have an in-depth analysis of the participants' creation and usage of multi-modal emoticons. We discuss the extracted underlying patterns and implications for future research and design below.

Implication 1: Purposefully leveraging the extra affordance of multi-modal emoticons. The variety of examples collected in the findings illustrate that by freely combining multi-modal elements, users were not only able to enhance or augment the

expression of static emoticons, but also create new meanings that are additional to, or different from the intended meaning of static emoticons. This suggests that multi-modal emoticons can offer extra communication affordance to existing emoticon systems, implying new research and design opportunities.

Studies have been conducted to explore the idea of *emoji affordance* on how users could build upon the visual or contextual properties of an emoji to enable richer meanings [52, 28, 1]. However, little has been done to understand and leverage the communication affordance of multi-modal emoticons. While serving as an early exploration, our work has surfaced that the multi-modal elements like animations and vibrotactile signals can facilitate users' communication beyond simply as subsidiary feedback to specific emoticons (which is how they have often been designed in current messaging apps). Instead, just like new meanings could emerge from the combination of multiple emoticons [28], our results suggest that new expressions could be created by combining static emoticons with different animations and vibrations. We thereby argue that future designs could purposefully help users leverage this new affordance to meaningfully expand their nonverbal communication vocabulary. To do so, a user-led approach is needed: researchers and designers could continually curate and analyze examples of users' creative expressions via multi-modal emoticons (e.g., Section 6.2.1), and intentionally facilitate similar use cases in design.

Implication 2: Designing for immersive and empathetic experience in online nonverbal communication. As uncovered by our empirical data, another promising opportunity of multi-modal emoticons is that they could create a certain atmosphere for a chat session, or make their responses feel more empathetic and attentive to each other. These new experiences in nonverbal communication were not supported by static emoticons. For instance, P15 always used a specific multi-modal emoticon to set the mood for joking, and P9 used multi-modal emoticons to render friendly, and relaxing vibes. Such immersive, and “*contagious*” ambience was, to a large extent, enabled by the animation effect which was continuously rendered as long as the emoticon is displayed on the chat screen. The future design could build upon this experiential quality and make online communication even more immersive: e.g., when a new multi-modal emoticon is sent, the ambient visual elements on the chat screen (e.g., chat bubbles or background) will react to it and adjust to the same atmosphere (e.g., a happy multi-modal emoticon will make the ambient elements look brighter), until the next emoticon shapes the ambience again.

As reasoned by the participants, the attentive, empathetic responses were enabled partially because they could create small variations of animations or vibrations in the multi-modal emoticons, instead of sending receptive or preset emoticons. Whereas, responding to others with the repeated static emoticon might feel in-polite and in-attentive. The empathetic experience was reported to be further enhanced when the two conversation

partners used similar animation or vibration to express similar feelings, which made them feel more “*connected*” and “*in sync*”. Inspired by this, the future design could explore enhancing such empathetic experience, e.g., by creating a resonance, or echo effect on the screen, when two users used similar multi-modal elements in chat.

Implication 3: Designing new animations and vibrotactile patterns based on body languages and everyday sounds. The great creativity exhibited by our participants in authoring multi-modal emoticons was supported by the building blocks: the sets of animations and vibrations prepared upon the open libraries of Kineticons [22] and VibViz [43]. Despite being comprehensive and well-designed, these libraries were intended to support a wide range of design purposes in graphical or haptic interfaces, rather than multi-modal emoticons. For this reason, as mentioned in Section 4.1, not the whole libraries could be meaningfully used in VibEmoji. Hence, a meaningful question in this emerging domain is: how we can design new building blocks, such as libraries of animation effects or vibrotactile patterns, to better nourish the creativity of both designers and everyday users of multi-modal emoticons.

Our findings from Section 6.2.2 could shed light on this question. Participants often associated their choice for vibrations with certain everyday sounds (e.g., knocking on the door, snoring, or tick-tock). Moreover, their choice for animations was often associated with bodily movements or body language (e.g., nodding, pushing, face-twitching). When there was a match between their communication intention and the association of the element, they would choose that element (e.g., P19 used a “*nodding animation*” to express agreement or hospitality; P3 used a “*snore vibration*” to express going to sleep). This interesting pattern suggests that new animations and vibrotactile signals could be designed to enable users’ richer associations with the two categories. For instance, future work could design new vibration elements based on a typical set of everyday sound effects. Or, designers could refer to a taxonomy of (bodily) nonverbal signals [50] to design new animations that could communicate more intentions of users (e.g., resembling certain eye behaviors or arm gestures, in addition to the head poses or full-body motions that they often mentioned).

Implication 4: balancing the design trade-off of automation versus autonomy. To support user-authoring multi-modal emoticons, VibEmoji features a non-aggressive recommendation approach, which prioritizes relevant options for users without limiting their free experimentation. As appreciated by the participants, such an approach leaves enough agency for them to create personalized expressions and avoids giving arbitrary and irrelevant suggestions due to misinterpreting the user’s intention or the context. However, we do recognize that a design tension exists, concerning the trade-offs between automation and autonomy. For instance, a number of prior works explored fully automating users’ entry of emoticons based on facial expressions [33, 12] or speech [23].

While automation often means low effort and high efficiency, our study revealed that in certain social communication scenarios, users value the autonomy and richness of self-expression more than efficiency. As reported earlier, while considering the recommendation principles, the users enjoyed being explorative and exploring different combinations. Therefore, an important consideration for future work is to better understand users' different needs for autonomy and automation in various social communication scenarios and carefully balance the two in design.

Limitations. Our work is not without limitations. First, VibEmoji included the same 50 (frequently used) emojis for each user to ensure that they had the same building blocks to begin with. Although abundant examples of their creative usage have been found, users could have more freedom if they could import new stickers to the system. A future field study can be conducted to enable each pair to co-design their libraries, in the form of the co-customization [20] for multi-modal emoticons. Second, participants in this field study were aged from 19 to 35; hence, how different age groups will use multi-modal emoticons and to what extent the findings might be applicable beyond the current user segment remain to be answered. It would be interesting in the future to involve older age groups or even elderly users, to understand their experiences. Additionally, the VibEmoji user interface might be further polished: its size could be made a bit smaller to display more historical chats above, and a search bar could be added to enable element selection by keyword. Lastly, VibEmoji currently employs a recommendation algorithm that only considers the usage frequency and patterns of one single user (i.e., the benefiting user). With more data collected, advanced recommendation techniques, such as collaborative filtering [32], can be achieved by mining common patterns across all the users and suggesting frequently-used multi-modal emoticon combinations of the crowd that is similar to the benefiting user.

Chapter 8

Conclusion

In this study, we set out to empirically understand multi-modal emoticons: an emerging phenomenon in current messaging applications which enhance static emoticons with multi-modal effects such as animations and vibrations. In doing so, we built VibEmoji, a user-authoring multi-modal emoticon system for mobile messaging. The design of VibEmoji extends beyond the current systems and allows us to extract new understandings about how users create, share, and experience multi-modal emoticons in daily communications, and how we could better support them by design. We evaluated VibEmoji with 20 participants in naturalistic settings over the period of four weeks. Findings were gathered to contextualize the creative usage and meaningful scenarios of multi-modal emoticons, as well as the users' detailed process of authoring. Based on the data, we generalize relevant design implications to inform future work in better supporting users' creation and communication of multi-modal emoticons.

References

- [1] Jessalyn Alvina, Chengcheng Qu, Joanna McGrenere, and Wendy E. Mackay. Moji-board: Generating parametric emojis with gesture keyboards. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery.
- [2] Apple. Use memoji on your iphone or ipad pro, 2021.
- [3] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2):77–101, 2006.
- [4] Scott Brave and Andrew Dahley. Intouch: A medium for haptic interpersonal communication. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, page 363–364, New York, NY, USA, 1997. Association for Computing Machinery.
- [5] Brian. Plato emoticons, revisited, 2012.
- [6] Keith Broni. Emoji use at all-time high, 2022.
- [7] Jessica R. Cauchard, Janette L. Cheng, Thomas Pietrzak, and James A. Landay. Activibe: Design and evaluation of vibrations for progress monitoring. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 3261–3271, New York, NY, USA, 2016. Association for Computing Machinery.
- [8] Zekun Chang, Tung D. Ta, Koya Narumi, Heeju Kim, Fuminori Okuya, Dongchi Li, Kunihiro Kato, Jie Qi, Yoshinobu Miyamoto, Kazuya Saito, and Yoshihiro Kawahara. *Kirigami Haptic Swatches: Design Methods for Cut-and-Fold Haptic Feedback Mechanisms*, page 1–12. Association for Computing Machinery, New York, NY, USA, 2020.

- [9] Zhenpeng Chen, Yanbin Cao, Huihan Yao, Xuan Lu, Xin Peng, Hong Mei, and Xu-anzhe Liu. Emoji-powered sentiment and emotion detection from software developers' communication data. *ACM Trans. Softw. Eng. Methodol.*, 30(2), January 2021.
- [10] Marcel Danesi. *The semiotics of emoji: The rise of visual language in the age of the internet*. Bloomsbury Publishing, London, UK, 2016.
- [11] Paul Ekman. Are there basic emotions? *Psychological Review*, 3(99):550–553, 1992.
- [12] Abdallah El Ali, Torben Wallbaum, Merlin Wasemann, Wilko Heuten, and Susanne CJ Boll. Face2emoji: Using facial emotional expressions to filter emojis. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, page 1577–1584, New York, NY, USA, 2017. Association for Computing Machinery.
- [13] Instagram Engineering. Emojineering part 1: Machine learning for emoji trends, 2015.
- [14] Mario J Enriquez and Karon E MacLean. The hapticon editor: a tool in support of haptic communication research. In *11th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2003. HAPTICS 2003. Proceedings.*, pages 356–362. IEEE, IEEE, 2003.
- [15] Facebook. Messenger, 2021.
- [16] Scott Fahlman. Smiley lore :-), Retrieved 2021.
- [17] Yunhe Feng, Zheng Lu, Wenjun Zhou, Zhibo Wang, and Qing Cao. New emoji requests from twitter users: When, where, why, and what we can do about them. *Trans. Soc. Comput.*, 3(2), apr 2020.
- [18] Google. Gboard, 2021.
- [19] Carla F Griggio, Joanna McGrenere, and Wendy E Mackay. Customizations and expression breakdowns in ecosystems of communication apps. *Proc. ACM Hum.-Comput. Interact.*, 3, 11 2019.
- [20] Carla F. Griggio, Arissa J. Sato, Wendy E. Mackay, and Koji Yatani. Mediating intimacy with dearboard: A co-customizable keyboard for everyday messaging. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2021. Association for Computing Machinery.

- [21] Loni Hagen, Mary Falling, Oleksandr Lisnichenko, AbdelRahim A. Elmadiany, Pankti Mehta, Muhammad Abdul-Mageed, Justin Costakis, and Thomas E. Keller. Emoji use in twitter white nationalism communication. In *Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing*, CSCW '19, page 201–205, New York, NY, USA, 2019. Association for Computing Machinery.
- [22] Chris Harrison, Gary Hsieh, Karl D.D. Willis, Jodi Forlizzi, and Scott E. Hudson. Kineticons: Using iconographic motion in graphical user interface design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, page 1999–2008, New York, NY, USA, 2011. Association for Computing Machinery.
- [23] Jiaxiong Hu, Qianyao Xu, Limin Paul Fu, and Yingqing Xu. Emojilization: An automated method for speech to emoji-labeled text. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI EA '19, page 1–6, New York, NY, USA, 2019. Association for Computing Machinery.
- [24] Huawei. Meetime, 2021.
- [25] Ali Israr, Siyan Zhao, and Oliver Schneider. Exploring embedded haptics for social networking and interactions. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, page 1899–1904, New York, NY, USA, 2015. Association for Computing Machinery.
- [26] Jialun "Aaron" Jiang, Jed R. Brubaker, and Casey Fiesler. Understanding diverse interpretations of animated gifs. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, page 1726–1732, New York, NY, USA, 2017. Association for Computing Machinery.
- [27] Ryan Kelly and Leon Watts. Characterising the inventive appropriation of emoji as relationally meaningful in mediated close personal relationships. *Experiences of technology appropriation: Unanticipated users, usage, circumstances, and design*, 2, 2015.
- [28] Sujay Khandekar, Joseph Higg, Yuanzhe Bian, Chae Won Ryu, Jerry O. Talton Iii, and Ranjitha Kumar. Opico: A study of emoji-first communication in a mobile social app. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, page 450–458, New York, NY, USA, 2019. Association for Computing Machinery.
- [29] Erin Kim and Oliver Schneider. *Defining Haptic Experience: Foundations for Understanding, Communicating, and Evaluating HX*, page 1–13. Association for Computing Machinery, New York, NY, USA, 2020.

- [30] Jihun Kim, Amitash ojha, Yongsik Jin, and Minho Lee. Pictogram generator from korean sentences using emoticon and saliency map. In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, HAI '15, page 259–262, New York, NY, USA, 2015. Association for Computing Machinery.
- [31] Philippe Kimura-Thollander and Neha Kumar. *Examining the "Global" Language of Emojis: Designing for Cultural Representation*, page 1–14. Association for Computing Machinery, New York, NY, USA, 2019.
- [32] Yehuda Koren and Robert Bell. *Advances in collaborative filtering*, pages 77–118. Springer, New York, NY, USA, 2015.
- [33] Miki Liu, Austin Wong, Ruhi Pudipeddi, Betty Hou, David Wang, and Gary Hsieh. Reactionbot: Exploring the effects of expression-triggered emoji in text messages. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), November 2018.
- [34] Paritosh Mittal, Kunal Aggarwal, Pragya Paramita Sahu, Vishal Vatsalya, Soumyajit Mitra, Vikrant Singh, Viswanath Veera, and Shankar M. Venkatesan. Photo-realistic emoticon generation using multi-modal input. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, IUI '20, page 254–258, New York, NY, USA, 2020. Association for Computing Machinery.
- [35] Joe Mullenbach, Craig Shultz, J. Edward Colgate, and Anne Marie Piper. Exploring affective communication through variable-friction surface haptics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 3963–3972, New York, NY, USA, 2014. Association for Computing Machinery.
- [36] Hideyuki Nakanishi, Kazuaki Tanaka, and Yuya Wada. Remote handshaking: Touch enhances video-mediated social telepresence. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, page 2143–2152, New York, NY, USA, 2014. Association for Computing Machinery.
- [37] Young-Woo Park and Tek-Jin Nam. Poke: A new way of sharing emotional touches during phone conversations. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, page 2859–2860, New York, NY, USA, 2013. Association for Computing Machinery.
- [38] Anand Rajaraman and Jeffrey David Ullman. *Data Mining*, page 1–17. Cambridge University Press, Cambridge, UK, 2011.

- [39] David Rodrigues, Marília Prada, Rui Gaspar, Margarida V Garrido, and Diniz Lopes. Lisbon emoji and emoticon database (leed): Norms for emoji and emoticons in seven evaluative dimensions. *Behavior research methods*, 50(1):392–405, 2018.
- [40] James A Russell and Lisa Feldman Barrett. Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. *Journal of personality and social psychology*, 76(5):805, 1999.
- [41] Oliver S. Schneider, Hasti Seifi, Salma Kashani, Matthew Chun, and Karon E. MacLean. Hapturk: Crowdsourcing affective ratings of vibrotactile icons. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, page 3248–3260, New York, NY, USA, 2016. Association for Computing Machinery.
- [42] Hasti Seifi, Matthew Chun, and Karon E. Maclean. Toward affective handles for tuning vibrations. *ACM Trans. Appl. Percept.*, 15(3), jul 2018.
- [43] Hasti Seifi, Kailun Zhang, and Karon E. MacLean. Vibviz: Organizing, visualizing and navigating vibration libraries. In *2015 IEEE World Haptics Conference (WHC)*, pages 254–259, New York, NY, USA, 2015. IEEE.
- [44] Andreas Sonderegger, Klaus Heyden, Alain Chavaillaz, and Juergen Sauer. *AniSAM & AniAvatar: Animated Visualizations of Affective States*, page 4828–4837. Association for Computing Machinery, New York, NY, USA, 2016.
- [45] Telegram. Telegram emoji platform, custom animated emoji packs, gifting telegram premium, and more, 2022.
- [46] Tencent. Wechat, 2021.
- [47] Gareth W. Tigwell, Benjamin M. Gorman, and Rachel Menzies. *Emoji Accessibility for Visually Impaired People*, page 1–14. Association for Computing Machinery, New York, NY, USA, 2020.
- [48] Maggie Tillman. What are animoji? how to create and use apple’s animated emoji, 2021.
- [49] Yuki Urabe, Rafal Rzepka, and Kenji Araki. Emoticon recommendation system for effective communication. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, ASONAM ’13, page 1460–1461, New York, NY, USA, 2013. Association for Computing Machinery.

- [50] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and vision computing*, 27(12):1743–1759, 2009.
- [51] Rongrong Wang, Francis Quek, Deborah Tatar, Keng Soon Teh, and Adrian Cheok. *Keep in Touch: Channel, Expectation and Experience*, page 139–148. Association for Computing Machinery, New York, NY, USA, 2012.
- [52] Sarah Wiseman and Sandy J. J. Gould. *Repurposing Emoji for Personalised Communication: Why 🍕 Means “I Love You”*, page 1–10. Association for Computing Machinery, New York, NY, USA, 2018.
- [53] Koji Yatani and Khai Nhut Truong. Semfeel: A user interface with semantic tactile feedback for mobile touch-screen devices. In *Proceedings of the 22nd Annual ACM Symposium on User Interface Software and Technology*, UIST ’09, page 111–120, New York, NY, USA, 2009. Association for Computing Machinery.
- [54] Amy X. Zhang, Michele Igo, Marc Facciotti, and David Karger. Using student annotated hashtags and emojis to collect nuanced affective states. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*, L@S ’17, page 319–322, New York, NY, USA, 2017. Association for Computing Machinery.
- [55] Mingrui Ray Zhang, Ruolin Wang, Xuhai Xu, Qisheng Li, Ather Sharif, and Jacob O. Wobbrock. Voicemoji: Emoji entry using voice for visually impaired people. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, New York, NY, USA, 2021. Association for Computing Machinery.
- [56] Zhuoming Zhang, Jessalyn Alvina, Robin Héron, Stéphane Safin, Françoise Détienne, and Eric Lecolinet. Touch without touching: Overcoming social distancing in semi-intimate relationships with sanstouch. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, New York, NY, USA, 2021. Association for Computing Machinery.
- [57] Zhuoming Zhang, Robin Héron, Eric Lecolinet, Françoise Detienne, and Stéphane Safin. Visualtouch: Enhancing affective touch communication with multi-modality stimulation. In *2019 International Conference on Multimodal Interaction*, ICMI ’19, page 114–123, New York, NY, USA, 2019. Association for Computing Machinery.
- [58] Rui Zhou, Jasmine Hentschel, and Neha Kumar. *Goodbye Text, Hello Emoji: Mobile Communication on WeChat in China*, page 748–759. Association for Computing Machinery, New York, NY, USA, 2017.