Chapter 8

Overall Discussion and Future Work

In the three individual studies in my thesis, I demonstrated that keystroke patterns are associated with different underlying intentions in a conversation, where those intentions may not be evident from visible word choice alone.

Study 1, by better identifying dialogue acts, lets us better understand what part of a conversation a user considers to be agreed upon as common ground, and which direction the user wants the dialogue to proceed in: a user might have a question about previous material. which means that they do accept the previous context as part of the common ground, or they may want to advance the conversation forward because both participants agree on what's been said (e.g. Convertino et al., 2008, but see their further clarification).

Study 2 uses keystrokes to better detect sentiment in utterances. While sophisticated methods exist that measure sentiment based on word choice alone, I show that adding keystroke information can improve these results. Because I am studying an interaction rather than a monologue, sentiment is also sensitive to changes from turn to turn. I show that keystrokes are also sensitive to these changes in sentiment, in addition to the sentiment of the utterance on its own. Another unique element of a dialogue is that a user forms an overall opinion of their partner. I show that this overall opinion also influences typing patterns, independently of the sentiment of that specific utterance.

Finally, Study 3 looks at how well keystrokes can predict the level of rapport established between a user and their partner. Further, study 3 asks which subsets of conversational data best predict rapport? I divided conversations chronologically into a first half and second half, divided a conversation by the task being accomplished, and randomly subset the conversational data. Using the findings from this study, rapport can be measured *during* a conversation, not just at the end.

The overall findings have not only theoretical implications for computer-mediated communication, but also practical implications for designing future CMC-related interfaces. If underlying intentions can be better identified, then a CMC system can utilize this information and either communicate it visually to a human partner or augment existing lexical information that a computer agent processes when conversing with a user. While the studies within my thesis were not designed to test this application directly, the research presented here is intended to act as a foundation for implementing these improvements to a CMC system.

CMC Theory

A main contribution of my thesis is the support it provides for a channel expansion theory (Carlson and Zmud, 1999), since the high levels of rapport achieved between anonymous users indicates that they were engaged in relatively rich conversation. Traditional theories of CMC such as media richness or bandwidth-limited theories posited that the capabilities of a medium are single-dimensional and determined *a priori*. A channel expansion theory, though, argues that as individuals gain more experience with a particular communication medium, the medium becomes richer for them (Carlson and Zmud, 1994). In other words, a medium's richness is perceptually determined by its users rather than being an intrinsic property of the medium itself. Walther (2011) situates this by saying that with experience, users learn how to encode and decode affective messages using a particular channel. In other words, the nature of media and their potentials are socially constructed (Fulk et al., 1987).

Kalman et al. (2013b) also points out that media richness theory does not explore the influence of the chronemic variables that are transmitted in lean media. In other words, media richness theory usually focuses on only the text of a text message and disregards elements such as the time between messages. But studies like my thesis show that a large amount of information is embedded in *how* language is produced, not just *what* language is produced. When this is not taken into consideration in analyzing a medium of communication, that analysis is necessarily incomplete and not an apt comparison between mediums.

In order to see how my experiments demonstrate this, it is instructive that Walther (2011) calls CMC users "cognitive and behavioral misers," who prefer to do a task using less effort than using more effort. As compared to face-to-face (FtF) communication, CMC is considered more effortful. If CMC is more effortful, though, then the overwhelmingly high rapport reported in study 3 shows that users are willing to incur greater effort, perhaps by making more use of the affordances of the experimental CMC platform (Clark and Brennan, 1991), in order to achieve the goal of the experiment, i.e. communicating recommendations and their rationales.

The high rapport ratings also support the filtered cues theory put forth in Walther and Parks (2002). In their critique of bandwidth-mediated theories, the researchers point out a confound in many previous studies that found FtF communication to be more expressive than CMC. While it has been well-established that CMC-based relationships take longer to develop than relationships built from in-person interactions, many previous studies used the same time limits for both mediums. For this reason I gave my participants 16 minutes to discuss a relatively specific topic (as opposed to completely spontaneous conversation or the more free-wheeling conversations in the Switchboard corpus (Godfrey et al., 1992)). It seems that this was an adequate amount of time to establish a high level of rapport, which helps to quantify the long*er* timeframes for relationship-building in CMC.

One question that my thesis brings up is whether certain theories of CMC are now outdated, or at least need heavy revision. At least in an environment such as the Prolific experimental platform, many participants have a very different relationship with CMC. Peer et al. (2022) reports that on some online behavioral research platforms, a majority of users use the platform as their primary

source of income and spend the majority of their day on the platform. These users have a very different relationship with technology in general and CMC specifically. The computer-mediated environment, for these users, is not one option among many, but is the sole option. In the Electronic Propiniquity Theory, the number of options matters, and users with only one option felt closer to the person they were communicating with (Walther, 2011). This could also explain the abundance of high rapport ratings among users in my experiments.

Nevertheless, as Walther (2011) reasons, "We can't keep up with new innovations, so we need theory and models that can (p. 754 in Scott 2009)." My experiments seem to add empirical evidence for existing theories of CMC that show that users are willing and able to expand the CMC medium to work for the purposes of their communication needs. However, it is possible that the shifting and more ubiquitous role of CMC will necessitate revisions to existing theories, as well.

Affective computing

My thesis adds to the rapidly growing field of *affective computing* (AC). This area of research recognizes that a computer agent needs to understand not only the (linguistic) content that a human is producing, but also recognize the emotions behind those words (Picard, 2000). Many methods for this recognition are either expensive or intrusive, such as galvanic skin responses, facial recognition, or voice analysis (Katerina and Nicolaos, 2018; Nahin et al., 2014, inter alia). Keystroke pattern analysis provides an unobtrusive and low-cost methodology for detecting user emotions.

Importantly for the many applications of my research, keystroke dynamics can capture multiple dimensions of emotion. As I demonstrated in Study 2, and has also been detected in prior research such as López-Carral et al. (2019), keystrokes can capture not only whether a user is experiencing a positive or negative emotion, but also the level of arousal, i.e. whether a user is mildly happy or ecstatic. This is one feature that sets keystroke analysis apart from a biometric such as facial expression recognition, where the latter often requires classification between a set of discrete basic emotions, rather than degrees of an emotion (Fasel and Luettin, 2003). The ability to identify

emotion on a continuous scale is equally, if not more important, than simply identifying *what* emotion a user is displaying.

Further, as shown in Study 3, keystrokes can also capture complex emotions such as rapport. Rapport is made up of many fundamental emotions, and so trying to directly ask about it or creating a single measure is difficult and less effective. Similar to Raj Prabhu et al. (2020), which measured multi-dimensional "conversation quality," I asked multiple questions after a conversation to construct a nuanced measure of rapport. Keystrokes were shown to be sensitive to this nuanced measure, demonstrating that keystrokes could also be used to detect complex, multi-dimensional emotions.

Ethical implications

Before discussing any practical applications of my research, it is imperative to discuss its ethical implications. Most crucially, keystrokes are a biometric, like a heart rate, a fingerprint or a voice pattern. Just as these are considered private or personal, keystrokes are also considered private or personal. Because of this my experiments required full final consent from participants, so that we had permission to collect and study their keystroke patterns. In addition, I received approval from Northwestern's IRB.

In addition to being a biometric for personal identification, keystroke patterns can also be used to identify demographics such as gender or education level (Brizan et al., 2015, inter alia). However most major internet browsers make it easy for developers to collect keystrokes (Acien et al., 2021). Thus, it is important to consider ethical implications when collecting keystroke information, as this information is both private and highly informative.

Any practical application discussed below could also be extended to more invasive and privacy-violating methods. The notion of identifying underlying intentions is very close to the idea of "Thought Police" (Orwell, 1949). As a recent example of the misuse of biometrics, many law enforcement agencies are using vocal analysis of 911 calls to determine if the caller is the actual perpetrator of the crime, despite the fact that this method has been debunked and multiple convictions

have been overturned (Murphy, 2022). Along these lines, it's not difficult to imagine a scenario akin to the movie *Minority Report* where law enforcement officers arrest people who they have determined *intend* to commit a crime, but before any crime has been committed. The "intentions" I identify in my thesis are somewhat ambiguous, and could be interpreted in many ways.

A dystopian use of keystroke analysis is already in place within the domain of employee surveillance. As more jobs become remote it becomes more difficult for employers to monitor employees. One method becoming increasingly popular for productivity surveillance is keystroke monitoring (Indiparambil, 2019; Tham and Holland, 2022). If an employee stops typing or is typing on a social media website, then the employer is notified. This has even been extended to using keystroke analysis to monitor employees' levels of motivation, and notify employers when employees are not motivated and likely not outputting high-quality work (Ball et al., 2021).

This also ties into the observation by some scholars that users sometimes *choose* to use text-based communication, even when other modalities are available, because they do not want their true intentions or mindset to be fully evident in a conversation. For example, Scissors and Gergle (2013) and Scissors et al. (2014) found that romantic couples would switch communication modalities to deescalate heightened emotions and avoid conflict. Similar, these studies found that users with low self-esteem find text-based CMC appealing because it mediates interactions, by lowering "face threat" and instead encouraging more "distancing" behavior.

Given findings such as those above, it is important to take into consideration that a user may not always *want* their complete mindset to be communicated via text-based CMC, e.g. by creating a visual representation of the emotions conveyed by typing patterns. Because of this, in any application of my findings in the real world, users would need the option to disable a visualization or anything in addition to the text appearing on the screen.

A more ethical approach to using keystrokes to monitor employee motivation has recently been undertaken at Microsoft Howe et al. (2022). Rather than reporting low motivation to a manager, a system that senses low motivation or stress suggests to that employee that they should take a break

and perhaps perform a stress-reduction exercise. In this way, mental states are not reported to a supervisor at all.

One advantage that keystroke analysis has above other forms of monitoring is that it can allow for anonymity while still extracting useful information about typing patterns from subsets of keystrokes. Monaco and Tappert (2016) presents a method for systematically obfuscating keystrokes so that a typist cannot be identified or impersonated. However, the anonymization does not negatively impact the typing experience: certain traits called "soft biometrics" can still be extracted from typing patterns, without even knowing the lexical content, and the keystroke obfuscation process was not noticeable or distracting to the user. The findings in Study 3 of my thesis also provide hope on this front since the findings demonstrate that rapport ratings can be extracted just using a subset of data rather than a full typing session. This could help to protect a user from sharing enough information that they can be personally identified.

Nonetheless, despite these advancements, privacy invasion is a very real concern when using keystroke analysis. This must be kept in mind when researchers or engineers use keystroke patterns to better understand a user.

Practical applications

The research in my thesis can be applied to any domain that utilizes text-based interaction. These include areas such as text-based telehealth, remote work on a Slack-like interface, or online chats with customer service. The common denominator in all of these scenarios is that it is critically important for one party to understand the mindset of the other party. Branigan et al. (2011) shows that we always try to infer the mindset in a conversation, whether a speaker believes they are interacting with another human or with a computer agent. However, inferring this mindset, or building a mental model of a partner, is difficult regardless of whether the partner is a human or computer agent (Gero et al., 2020; Yan et al., 2020).

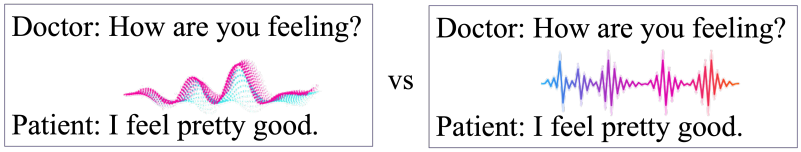

Figure 8.1
A toy example of different typing patterns visualized with the same lexical content

Typing data can tell us not just *what* is said, but *how* it's said. The style of delivery can be just as important as lexical content when trying to understand underlying mindsets of a language producer (Cowan et al., 2015). For instance, it has been well established that lexical alignment between speakers is a sign of positive affect between the partners. However, typing patterns are partially reflective of cognitive load (Brizan et al., 2015), and so typing patterns could inform observers about the level of cognitive effort underlying lexical alignment. If cognitive effort is significantly different across instances of lexical alignment, then perhaps not all lexical alignment signifies the same relationship quality between partners.

Thus, as an immediate extension of my thesis research, we can use important typing features to create a visual representation of the typing style behind the words transmitted, so as to inform a partner about the mindset behind the words. As an example, Figure 8.1 shows a fictional interaction between a doctor and patient. In this scenario, the colored waves accompanying the text are a visual representation of the patient's typing patterns, where the smooth wave on the left represents a consistent or consistent typing style, while the jagged wave on the right represents an inconsistent, hesitant typing style. Although the lexical output is identical, it is clear that the typing patterns that produced them are very different. As a result, a doctor should interpret "pretty good" very differently in each case.

In reality, though, a visual encoding schema such as that in Figure 8.1 would need significant refinement before being put into production. While some visual encodings might inherently convey meaning, these can be difficult to consistently produce (Iliinsky and Steele, 2011). More likely, users of an application that visualizes typing patterns would need to be explicitly instructed as to what each visualization means, as well as what that typing pattern means about the typist's mindset.

A CMC system that is aware of keystroke patterns could also assist human-to-human CMC by mediating a conversation. For example, Gergle et al. (2004) shows how the amount of chat history displayed affects the collaborative success of an interaction. If a computer agent moderating a conversation used typing patterns to informs its decisions, the computer agent could dynamically decide how much chat history to display in order to improve collaboration. If the computer agent determines via typing patterns that an utterance was produced under duress or during a moment of anger, than it may be conducive to collaboration if this utterance is quickly removed from the visible chat history. Conversely, if a backward-facing dialogue act was produced, asking for additional clarification, then a computer moderator could keep that utterance in the conversation's visible history for longer. Since study 2 of my thesis showed different typing patterns in forward- vs backward-facing dialogue acts, this is an area where typing information could be utilized.

Pommeranz et al. (2012) shows that people are willing to spend more effort if the feedback mechanism enables them to be more expressive. A feedback mechanism that takes into account production patterns, such as typing patterns signaling sentiment or opinions, would allow a user to be more expressive. Thus they may put more effort into their interaction. If this took place, it would also lend support to a channel expansion theory (Walther, 2011; Walther and Parks, 2002).

Implicit Prosody Hypothesis

While my thesis is not intended to prove or disprove any specific cognitive theories, it does use cognitive science as a basis for its investigations. Specifically, the features used in all of my studies are based on features of spoken prosody, in order to test the Implicit Prosody Hypothesis (IPH Fodor, 2002a).

Spoken prosody has proved to be useful in many HCI settings such as trust development (Beňuš et al., 2018), communication efficiency and emotional engagement (Suzuki and Katagiri, 2007), naturalness in an interaction (Bell et al., 2003), and mental models of a computer conversational

partner (Cowan et al., 2015). Spoken prosody has also been shown to be useful in DA classification, sentiment analysis, and rapport detection (see previous Related Work chapters).

Prior studies have found pause locations in keystrokes to be similar to pause locations in spoken language. Plank (2016) and Goodkind and Rosenberg (2015) found that typists pause for longer at the boundaries of syntactic units, which is also a feature of speaking. In their research, Plank (2016) was actually able to use keystroke timing as a crude syntactic parser.

In my own studies, keystroke features based on prosodic features improved the accuracy of identifying underlying motivations and mindsets of participants engaging in dialogue, which are also commonly signaled in spoken interaction using speech prosody. While this does not constitute incontrovertible proof that a user is utilizing the exact same prosodic cognitive pathways when typing as they are when speaking, it does provide a possible association and an intriguing foundation for future research.

Since spoken prosody is at least partially determined by a speaker's relationship with the audience they are addressing (audience design, Horton, 2017), typing patterns could also be informative about a typist's relationship with their audience. Future research should look for more direct evidence of parallels, which would open the door for using better-studied features of speech prosody to be co-opted in keystroke dynamics.

Typing metrics as an independent and dependent variable

In my respective experiments, I used keystrokes both as a response variable and a predictor. As I explain below, this speaks to the distinctiveness of keystroke patterns in different situation, as well as their consistency within a situation. Using the various definitions of a "non-verbal cue" provided in Kalman (2007), it seems that the consistently different properties of keystrokes that I have identified would make them a proper cue.

In Study 1, Exp 1a treated a dialogue act binary as the dependent variable with a set of keystroke metrics as independent predictors, but Exp 1b treated a single keystroke metric as a dependent

variable with dialogue acts as a single independent predictor with multiple levels. In Study 2, Exp 2a treated a sentiment level binary as the dependent variable with a set of keystroke metrics as independent predictors, but Exp 2b treated a single keystroke metric as a dependent variable with sentiment level and opinion ratings, respectively, as independent predictors. Given this, a natural question to ask is why keystrokes are both predictors as well as response variables.

In Studies 1 and 2, when a set of keystroke metrics were independent predictors, the question at hand was whether keystrokes, collectively, could be used to differentiate between a binary distinction. The question was not whether a single typing pattern could be used as a differentiator, as it has been well-established that a mix of keystroke metrics provide for a better overall model. In these experiments, the set of keystroke metrics had to produce unique values for each level of the dependent variable.

On the other hand, when a single keystroke metric was used as a response variable, the question at hand was whether each level of the predictor(s) had a robust timing signature for that keystroke metric or resulted in a consistent change. The signature at each level did not need to be unique, but rather needed to be consistent.

Limitations and shortcomings

The nature of my study design, data collection methods, and recruited population imposed limitations on how generalizable the results reported in my thesis will be. This is not to say that my study was flawed; rather, every study sets boundaries for specificity and these impose limitations on the scope of the findings.

The limitations imposed by the study design can be broken into three facets: 1) the experimental interface, 2) the experimental prompts, and 3) the recruited population.

As seen in Figure 4.7, the experimental interface used a very generic aesthetic, which looked very dissimilar to a modern chat client such as Slack or Google Messenger. Studies such as Branigan et al. (2011) have shown that humans interact differently with a computer agent depending on

whether the aesthetics of the interface make the agent look sophisticated. It is possible that this extends to human-human conversations. Since the interface did not look anything like a chat interface on a participant's phone or computer, the participant would be more consciously aware of the interface and how different it looked from the interfaces they are used to. As such, the typing patterns collected would not be as naturalistic as the typing patterns collected when in Slack or on an iPhone.

In addition, I disabled autocorrect, autocompletion, use of emojis, and the ability to change aspects of the font. Users of modern chat interfaced are likely very used to autocorrect and autocompletion, and so imposing the necessity of typing out an entire word and being more self-conscious of correct spelling also made for a less naturalistic experience. In addition, studies such as Liebman and Gergle (2016b) show that emoticon use is used in text-based communication to mediate interpersonal affinity. By depriving my subjects of the ability to use emojis or convert emoticons to emojis, I also deprived them of a tool used to convey emotion in a text-based environment.

My experimental prompts also provide limitations for my collected data. My goal was to strike a balance between a tightly-controlled but less naturalistic conversation and completely free-wheeling conversations that were different lengths and contained completely different content. In order to achieve this goal, my prompts included the role that each participant would play ("receiver" versus "provider" of recommendations), the general genre of movies/TV shows to discuss, and some pointers about what the conversation should look like, i.e. short messages akin to a text message conversation with a friend.

Very few real-world interactions would have all of the constraints I imposed. Certainly, a typing analysis used in a real-life setting would need to be able to accommodate very different content, and messages of very different lengths. In addition, a real-life scenario would involve implicit role assignment, but these assignments would not be explicitly delineated. For example, if a user logs onto a telehealth platform for medical advice, they would implicitly take on the role of recipient. However, during a conversation these roles will naturally fluctuate, e.g. the patient *providing*

medical history to their doctor. As a result, while Study 3 of my thesis found a high degree of accuracy using the Receiver subset, in the real world it would not be trivial to partition this data.

An additional limitation was the recruited population. Because I recruited exclusively from a crowdsourcing platform, where users were required to have a fair amount of experience with online experiments, it is safe to assume that the participants had familiarity with computer interfaces, and typed more fluidly on average. However, an application deployed in the real world will have uses of varying typing competence. As such, it would need to accommodate a naturally slow or inaccurate typist, but understand that this is not necessarily indicative of high cognitive load or hesitation.

Similarly, one data collection methodological issue was that participants used different types of computers alongside different internet connection speeds. If a typing application was deployed in the wild, the developers would need to account for different keystroke latencies across devices/connections. Whether these timing differences are perceptible by humans should be studied in the future. However, if fine-grained keystroke timing differences are necessary for an analysis, then this needs to be taken into account.

A major experiment methodological limitation was that the conversations in my study lacked a tangible goal. While exchanging recommendations was a "goal," there was no objective measure of success. For example, in Kalman et al. (2010) participants played a trading game where success depended on collaboration and the outcome of the game partially determined compensation. Because of this, participants were motivated to collaborate, and the number of successful trades was used to operationalize collaborative success. In my studies, there was no way to measure a successful outcome and participants had no enticements to collaborate.

Despite this, while it is not an objective measure, the high rapport ratings in study 3 show that participants were willing to invest effort into making high-quality movie and TV show recommendations. This occurred despite that my experiment did not involve additional incentives for good recommendations. This speaks to features of CMC such as *fluidity*, where ease of communication induces better communication; Faraj et al. (2011) cites as a reason for free labor in online communities such as Wikipedia (Rafaeli and Ariel, 2008). Because a text-based chat interface provides

few barriers to knowledge sharing, participants are more willing to contribute effort to providing high-quality recommendations.

In addition, my experimental methodology was perhaps too naturalistic and not controlled enough to get stronger experimental results and establish a more firm foundation. As keystroke analysis of dialogue is a relatively unstudied area, it is possible that a semi-spontaneous conversation did not provide enough controls for a foundational study. That being said, the full dataset, called the Keystrokes in Dialogue (KiD) Corpus, is available at https://github.com/angoodkind/KiDcorpus. This data should be of great value to the larger HCI community and allow for the foundational studies in my thesis to be expanded upon.

Conclusion

Despite the shortcomings of the work in my thesis, the research does provide important contributions to the fields of HCI and cognitive science. Typing, especially in interactions, is a fascinating modality to explore because it combines together elements of language production, language comprehension, and a dyadic relationship. For example, when a message is being sent its composition is only viewable by the producer, but its final product is shared with a partner. Thus, typing production in text-based messages allows us to understand an interaction both as an internal process as well as a shared process. Findings from my thesis shed light on theories of why we communicate in the way we do, and how this effect is received by those we interact with.

Finally, the vast majority of typing studies have been performed on typing in isolation. My thesis research studies typing, though, as an interactive process, where "language is used for doing things (Clark, 1996, p. 3). My hope is that my thesis research, in turn, can be used to improve human communication and allow us to collaborate more successfully via text-based computer-mediated communication.