Practical Challenges for Investigating Abbreviation Strategies

Elisa Kreiss*
ekreiss@stanford.edu
Stanford University
Stanford, CA, USA

Subhashini Venugopalan
vsubhashini@google.com
Google Research
Mountain View, CA, USA

Shaun Kane
shaunkane@google.com
Google Research
Boulder, CO, USA

Meredith Ringel Morris
merrie@google.com
Google Research
Seattle, WA, USA

ABSTRACT

Saying more while typing less is the ideal we strive towards when designing assistive writing technology that can minimize effort. Complementary to efforts on predictive completions is the idea to use a drastically abbreviated version of an intended message, which can then be reconstructed using Language Models. This paper highlights the challenges that arise from investigating what makes an abbreviation scheme promising for a potential application. We hope that this can provide a guide for designing studies which consequently allow for fundamental insights on efficient and goal-driven abbreviation strategies.

CCS CONCEPTS

• **Human-centered computing** → **HCI design and evaluation methods**; **Accessibility**.

KEYWORDS

abbreviation strategies, assistive writing, cognitive load

1 INTRODUCTION

Many technological advancements are guided by the desire to speed up daily routines. Communication is one such routine and the focus of our work. Most face-to-face conversations are rapid exchanges of information (ca. 130 words per minute), and when they have to be done via writing, they can be painstakingly slow. A natural strategy to speed up writing is the use of abbreviations, i.e., cutting down on the number of letters and words, which can be inferred back by the reader. While usually the writer is tasked with the challenging problem of abbreviating text to maximize the accuracy of the intended message by the recipient, recent advancements in Language Models (LMs) enable us to consider writing assistants that can reason about the abbreviated input to infer and expand the message before it is sent to the recipient. For instance, [4] provide a language model that can reliably infer a full sentence simply based on the first letter of each word, resulting in a keystroke reduction of up to 77%. These models can be integrated into text entry interfaces (e.g., [3, 7]) and can significantly reduce the number of keystrokes necessary while still allowing the recovery of the message. This doesn't just have implications for assistive writing but assistive communication more generally. People who can only communicate, for instance, through gazetyping (i.e., using a keyboard that is controlled through eye movement) can only communicate at rates of about 8-10 words per minute and each additionally saved keystroke increases the ability to participate in real-time conversations [9].

Building an understanding of what intuitive and efficient abbreviation schemes might look like needs to be at the center of this new avenue for facilitating keystroke minimization, but research that aims to identify best practices face several challenges that we outline here. We specifically focus on the case where the messages are intended for people other than ourselves, since idiosyncratic note-taking strategies pose a distinct problem. We argue that studies on abbreviation schemes in this domain need to optimize for four goals: two pertaining to the creation of abbreviated input by a user, and two regarding the output reconstruction, e.g., by a decoding system. For the input, an ideal abbreviation strategy (1) minimizes input time (e.g., reduces keystrokes, increases WPM, minimizes need for error corrections) and (2) minimizes the cognitive load required to enter the abbreviation (e.g., abbreviating with the initial letter of each word is cognitively much less demanding than abbreviating with the second letter). The output should (3) successfully represent the communicative intent of the message (message content) and (4) accurately portray the user's tone and stylistic details. In this work, we identify these challenges and present two studies that highlight complexities for research on abbreviation strategies that can inform applications.

2 DESIGNING ABBREVIATION STUDIES

Building models that can take abbreviated input and then extend it needs to be rooted in a complex investigation on user needs and usefulness. Studying the use of abbreviations, however, is a challenging feat. Here, we lay out those challenges and show in two case studies how the experimental choices affect the study outcomes.

2.1 Input: Speed & Ease

A straightforward measure of an abbreviation scheme's usefulness is the number of keystroke savings for each input [10]: the less a user has to type, the better. However, the cognitive effort involved to generate such an abbreviation is just as important to determine whether a proposed method can be successful in practice.

Investigating what makes a good abbreviation strategy can be hypothesis-guided or exploratory. In a *hypothesis-guided* study, participants are provided with specific rules for how to abbreviate (e.g., type only the first letter of every word) and the experiment focuses on the success and failure cases of a proposed abbreviation scheme. The advantage of this study design is that it allows for

*This work was conducted as part of a Google internship.

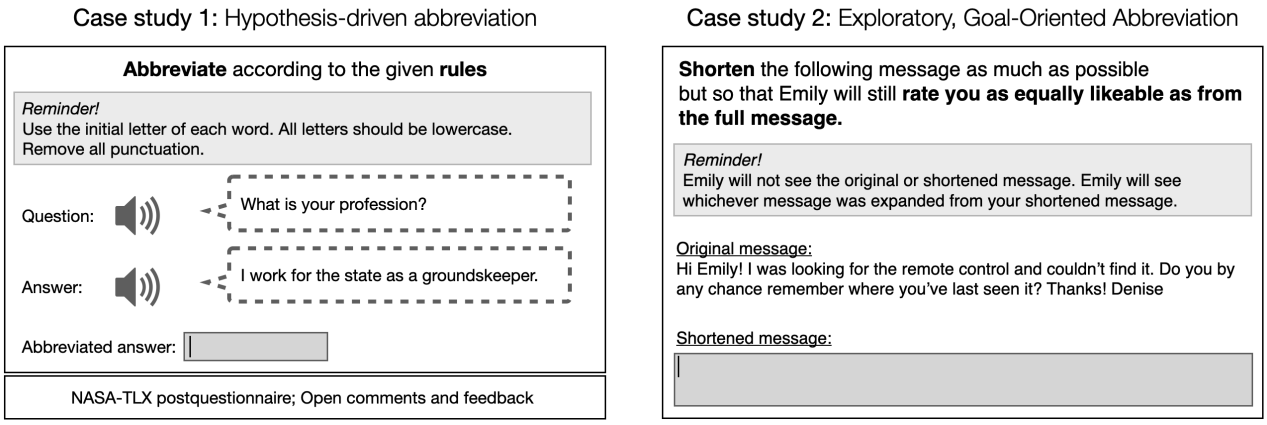

Figure 1: Overview of the two case study designs. In case study 1, participants were asked to apply a given abbreviation strategy to auditory stimuli. The goal of the participants was to abbreviate the sentences with minimal mistakes and as quickly as possible. In case study 2, participants developed their own abbreviation strategy and shorten a text visually given to them. Their goal was to abbreviate as much as possible, while still allowing the receiver of the message to, for instance, rate the sender of the message as equally likable.

data that's easy to analyze and provides detailed insights into the challenges associated with a proposed scheme. However, the insights are restricted to the predetermined scheme. In an *exploratory* study, no specific abbreviation scheme is provided. Instead, the task is formalized, for instance to minimize keystrokes under certain constraints. The analysis focuses on identifying which strategies users intuitively employ. This data allows rich insights on intuitive abbreviation strategies with minimal effects of prior preconceptions by the designers and experimenters. However, the experimental design of such studies can be challenging since it requires that participants reason about a system that will expand the shortened message which they likely have incomplete knowledge about (what are its overall capabilities and where will it fail).

Taken together, both exploratory and hypothesis-guided studies provide valuable insights and should be used to inform each other. Here, we present two case studies for two distinct experimental paradigms and discuss implications of the results.

2.2 Output: Accuracy of Message & Tone

An abbreviation scheme can only be useful as long as it still allows a faithful recovery of the original message. The most straightforward approach is to design models that can recreate the entirety of the original message character-by-character. The advantage is that it guarantees to preserve the intended message and meta characteristics like the message's tone and style. However, in many circumstances it may be the case that an exact replication is less important than getting a core message across with fewer keystrokes. We investigate this intuition and find evidence that users vary their abbreviation strategies based on their own preferences and situations. These results suggest the need for a more context-sensitive approach for evaluating the usefulness of abbreviation systems.

3 CASE STUDY 1: HYPOTHESIS-DRIVEN ABBREVIATION

Figure 1 provides an overview of the design for both case studies. In the first study, we provided a specific abbreviation scheme to investigate its challenges.

Design: We tested a very simple but effective abbreviation scheme where each word was shortened to its initial letter and punctuation was removed (inspired by [4]). After familiarizing themselves with the rules, participants proceeded to the main trials. In each trial, participants listened to, and then abbreviated spoken sentences of varying length, generated by a Text-to-Speech system (WaveNet, [8]). To avoid misperceptions of the text, all responses to be abbreviated were contextualized with a question, where Voice A asked a question and Voice B responded, which was the part that needed to be abbreviated. We used auditory as opposed to written stimuli to mimic a more natural setup where abbreviations are not visually copied. We note, however, that an auditory signal isn't quite the most natural open response style (i.e., it doesn't require the process of thinking about a response while abbreviating) but it allows for better control over the analysis since all users abbreviate the same text. The data participants abbreviated contained sentences that are unambiguously identifiable to type [1], from natural phone conversations (SwDA, [6]), and carefully designed sentences that minimally vary between each other to investigate patterns of interest. The effectiveness of the scheme was established through error and reaction time analyses, and open comments. Cognitive load was measured with the NASA TLX scale [5].

Results: We recruited 38 participants on Amazon's Mechanical Turk. The analysis of the data uncovered edge cases where the instructions were underspecified (how to treat compound nouns and contractions), which were reflected in the open comments and

disagreements between responses. We further found that the auditory signal significantly affects how the abbreviation scheme is applied. Specifically, what constitutes a word depends on the auditory signal: if the audio used "you're", participants were more likely to choose the abbreviation "y", but participants abbreviated "you are" as "ya". This pattern generalized over a variety of contractions.

4 CASE STUDY 2: EXPLORATORY, GOAL-ORIENTED ABBREVIATION

In the second study, we didn't provide an abbreviation scheme but instead a communicative goal that the message needed to achieve.

Design: Participants were instructed to abbreviate a message as much as possible while also still letting the receiver of the message fulfill a task. Participants were allowed to delete characters and words but not add words or rewrite (inspired by [2]). These constraints were necessary to allow for quantitative comparisons between messages from different participants. In each trial, participants were given a message that needed to be shortened as much as possible, together with a communicative goal that their version of the message should still manage to achieve. This goal could be either factual (e.g., the receiver should know what to bring from the store), or one that communicates some meta information about the speaker (e.g., the receiver should judge the message writer as equally intelligent as from the full message).

Results: We recruited 74 participants on Amazon's Mechanical Turk, and make the following two main observations. Firstly, with the goal to communicate factual information, messages were significantly shorter than when users were trying to preserve a characteristic trait. This suggests that depending on the goal, users employ distinct strategies when it comes to abbreviating content, which posits a new challenge for evaluating the practical usefulness of an abbreviation expansion model. Secondly, in this particular experiment, we allowed participants to delete content since the only restriction was the fulfillment of the goal. Consequently, the abbreviation strategies in this design were primarily content-cutting and less focused on removing characters. Understanding which incentives might encourage a trade-off between these strategies would be an interesting avenue for future work.

5 CONCLUSION

Allowing users to minimize the number of keystrokes necessary to communicate an intended message is a desirable goal in domains such as assistive writing and assistive communication. However, determining what doesn't only enable the maximum number of keystroke savings, but also aligns with minimized cognitive effort and goal-driven effectiveness is challenging to establish. In this paper, we provide a perspective on the challenges that come with empirical studies of abbreviation strategies. We provide case studies for both an exploratory and hypothesis-driven abbreviation scheme experiment, which provide complementary insights to people's abbreviation strategies. We hope that this work motivates the development of tools and platforms that allow more holistic insights on what makes a good abbreviation scheme for users, depending on their specific contextual goals and needs.

REFERENCES

[1] Jacob Abbott, Jofish Kaye, and James Clawson. 2022. Identifying an aurally distinct phrase set for text entry techniques. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.

[2] Jiban Adhikary, Jamie Berger, and Keith Vertanen. 2021. Accelerating Text Communication via Abbreviated Sentence Input. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics, Online, 6574–6588. https://doi.org/10.18653/v1/2021.acl-long.514

[3] Shanqing Cai, Subhashini Venugopalan, Katrin Tomanek, Shaun Kane, Meredith Ringel Morris, Richard Jonathan Noel Cave, Bob MacDonald, Jon Campbell, Blair Casey, Emily Kornman, et al. 2023. SpeakFaster Observer: Long-Term Instrumentation of Eye-Gaze Typing for Measuring AAC Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. https://research.google/pubs/pub52025/

[4] Shanqing Cai, Subhashini Venugopalan, Katrin Tomanek, Ajit Narayanan, Meredith Morris, and Michael Brenner. 2022. Context-Aware Abbreviation Expansion Using Large Language Models. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Association for Computational Linguistics, Seattle, United States, 1261–1275. https://doi.org/10.18653/v1/2022.naacl-main.91

[5] Alex Cao, Keshav K Chintamani, Abhilash K Pandya, and R Darin Ellis. 2009. NASA TLX: Software for assessing subjective mental workload. *Behavior research methods* 41 (2009), 113–117.

[6] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics* 26, 3 (2000), 339–373.

[7] Stephanie Valencia, Richard Cave, Krystal Kallarackal, Katie Seaver, Michael Terry, and Shaun K. Kane. 2023. "The less I type, the better": How AI Language Models can Enhance or Impede Communication for AAC Users. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*.

[8] Aäron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alexander Graves, Nal Kalchbrenner, Andrew Senior, and Koray Kavukcuoglu. 2016. WaveNet: A Generative Model for Raw Audio. In *Arxiv*. https://arxiv.org/abs/1609.03499

[9] Annalu Waller. 2019. Telling tales: unlocking the potential of AAC technologies. *International journal of language & communication disorders* 54, 2 (2019), 159–169.

[10] Jacob O Wobbrock. 2007. Measures of text entry performance. *Text entry systems: Mobility, accessibility, universality* (2007), 47–74.