

Using less keystrokes to achieve high top-1 accuracy in Chinese clinical text entry

DIGITAL HEALTH
Volume 9: 1–9
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231179027
journals.sagepub.com/home/dhj



Tao Li, Lei Yu, Liang Zhou and Panzhang Wang 

Abstract

Background: As a routine task, physicians spend substantial time and keystrokes on text entry. Documentation burden is increasingly associated with physician burnout. Predicting at top-1 with less keystrokes (TLKs) is a hot topic for smart text entry. In Western countries, contextual autocomplete is deployed to alleviate the burden. Chinese text entry is intercepted by input method engines (IMEs), which cut off suggestions from electronic health records (EHRs).

Objective: To explore a user-friendly approach to make text entry easier and faster for Chinese physicians.

Methods: Physicians were shadowed to uncover the real-word input behaviors. System logs were collected for behavior validation and then used for context-based learning. An in-line web-based popup layer was proposed to hold the best suggestion from EHRs. Keystrokes per character and TLK rate were evaluated quantitatively. Questionnaires were used for qualitative assessment. Nine hundred fifty-two physicians were enrolled in a field testing.

Results: 14 facilitators and 17 barriers related to IMEs were identified after shadowing. With system logs, physicians tended to split long words into short units, which were 1–4 in length. 81.7% of these units were disyllables. Compared to the control group, the intervention group improved TLK rate by 40.3% ($p < .0001$), and reduced keystrokes per character by 48.3% ($p < .0001$). Survey results also promised positive feedback from physicians.

Conclusions: Keystroke burden and frequent choice reaction time challenge Chinese physicians for text entry. The proposed system demonstrates an approach to alleviate the burden. Contextual information is easily retrieved and it further helps improve the top-1 accuracy, with a smaller number of keystrokes. While positive feedback is received, it promises a benefit to protect patient privacy.

Keywords

Electronic health records, input method engines, keystroke burden, reaction time, burnout

Submission date: 24 October 2022; Acceptance date: 12 May 2023

Introduction

Overloaded clinical documentation is a leading factor for physician burnout.^{1,2} Every day physicians click thousands of times especially on text entries in clinical notes.³ The time spent on direct patient care is reduced by half.^{4,5} Physicians have to act as a skilled typist in order to spend more time with patients.⁶ Scribe-based solutions are sought to alleviate the burden,^{7–10} implemented through either human or artificial intelligence. However, barriers persist such as extended labor and suboptimal accuracy.⁸ Self-administered manual entry is still a preferred modality.¹¹

Physicians often used abbreviations for saving keystrokes in letter-by-letter input,^{12–14} but it leads to unstructured content, which hinders understandability for patients, other physicians, and machines.^{15–18} Recent

Information Technology Department, Shanghai Sixth People's Hospital, Shanghai, China

Corresponding author:

Panzhang Wang, Information Technology Department, Shanghai Sixth People's Hospital, No.600, Yishan Road, Shanghai, 200233, China.
Email: pz_wang@hotmail.com



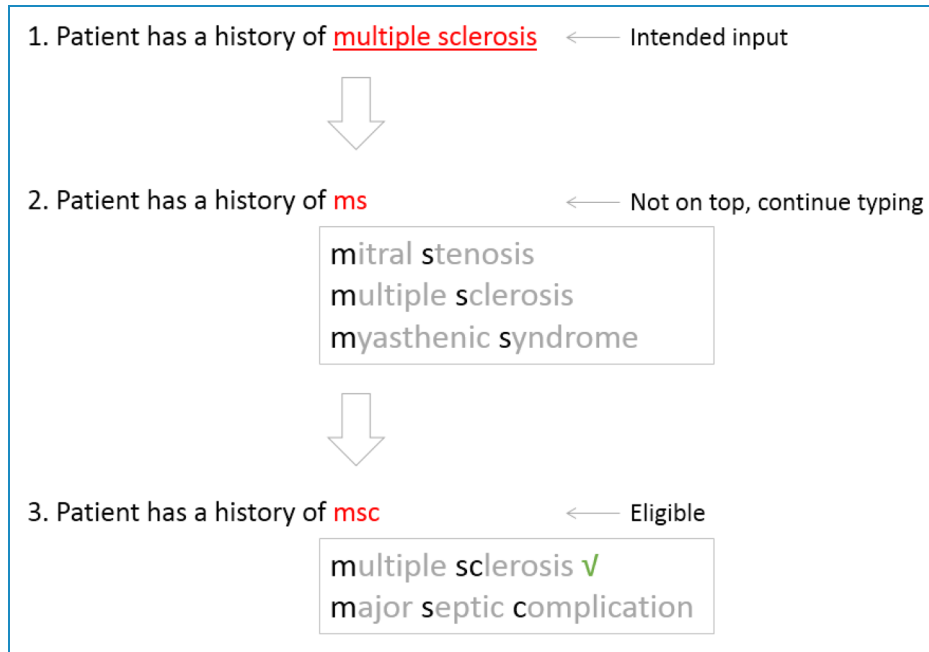


Figure 1. Suggest a word at the top position with less keystrokes.

improvements focus on using abbreviations transitorily through autocomplete.^{19–21} It involves several interactive steps: (1) physicians cognitively generate an acronym for a structured term, (2) physicians type it into notes, (3) a popup menu is displayed in-line with top-5 suggestions from electronic health records (EHRs), (4) physicians scan the list, and (5) physicians navigate and select a word. Suggestions would update for each entered letter. When the intended word is located at top-1, it can be selected by clicking Enter; otherwise, physicians have to navigate using the Up/Down and then press Enter.

Unlike direct input entered letter-by-letter, the suggestion-based input (SBI) uses an abbreviated input as a query parameter to fetch a list of normalized candidates. As shown in Figure 1, the suggestion list is updated for each letter entered into the parameter area. Normally, users will not stop to disambiguate each entered letter, as continuous typing within a single learned chunk is preferred by most typists.²² When the intended word is not in the list, the parameter is then appended or modified. Before the wanted word is found out, it might need to go through several turns of trying. After that, making a selection from the list needs some reaction time. According to Fitts' Law,²³ the reaction time spent on the top-1 item is the shortest. The top-1 candidate is popularly bound with a shortcut (e.g. the Enter), which makes the selection more touch-typing. Compared to direct input, SBI improves efficiency by using less keystrokes. A selection on the top-1 with less keystrokes (TLKs) seems a jackpot to the typist.

Chinese text entry is born to be a SBI, with the help of input method engines (IMEs).²⁴ Pinyin is the official

romanization system, commonly used as a phonetic system for writing.²⁵ Like autocomplete, IMEs consume Pinyin and output a list of Chinese candidates. Averagely, each character's Pinyin has 4.2 Latin letters,²⁶ which is either a vowel or a consonant–vowel combination. Mandarin Chinese uses only around 410 distinct syllables to encode more than 6000 characters.²⁷ Thus, using Pinyin causes ambiguities. Multiple characters form a linguistic word, and 64% of words have only two characters.²⁷ For efficiency, word-based input is widely used.^{28,29} Pinyin could also be abbreviated. For example, the full Pinyin for the Chinese word of “asthma” is “qi chuan.” It can be shortened to “qc,” “qch,” “qic,” “qich,” or “qchuan.” Basically, the shorter the sequence is, the more ambiguities it has.

Like autocomplete, IMEs also used the Up/Down for candidate navigation. However, IMEs use the Spacebar for selection. The Enter is reserved for direct input of Western words. Besides, IMEs use numeric keys for direct selection, although it is not friendly to touch-typing.²⁶ Unless the target word is located at top-1, about 52% of the input time is used for choice reaction and candidate selection.²⁶ Apparently, TLK is also important for Chinese physicians. However, as a standalone third-party tool, IMEs rank candidates, ignoring the possible suggestions from EHRs.

There are several popular IMEs on the market, such as Sogou Pinyin,³⁰ Baidu IME,³¹ and Google Pinyin.³² In September 2018, the vendor of Sogou Pinyin³⁰ launched a specific version for Chinese physicians,³³ shipped with 90,047 medical terms. The commercial tools are all free of charge, and physicians are allowed to use their favorite ones.³⁴ In 2019, Yang et al. developed a tool to evaluate

the domain-specific version against other general IMEs for medical writing.³⁴ Surprisingly, it failed to outperform the others. Although the evaluation is limited to full Pinyin in a lab, it suggests that commercial IMEs are converging in terms of input performance. For years, vendors' endeavors are limited to the kernel including a built-in vocabulary.^{35–38} Increasing the underlying word length seems friendly to TLK. However, to our knowledge, it receives limited attention in the clinical setting.

As an initial step to improving text entry for medical records, we shadowed physicians in the clinical setting to understand their input behaviors in the time-intensive environment. We designed a context model for note-taking and used system logs to collect contextual entries. With the learned knowledge from logs, we developed and evaluated a tool to promote TLK for Chinese note-taking. Our overall goal was to explore a user-friendly approach to make text entry easier and faster for Chinese physicians. It is the first study to use server-based rules to enhance client input in an IME-based setting, to our knowledge.

Methods

Study setting

This study was approved by the institutional review board of Shanghai Sixth People's Hospital. It was completed in three steps. First, shadowing was used to collect barriers and input behaviors associated with IME-based clinical documentation from the frontline. Second, a web-based tool was developed to enhance input behavior. Finally, a controlled study was conducted for quantitative and qualitative assessment.

Workspace shadowing

To mitigate disparity between self-reported input behavior and the actual observable patterns, the team conducted workspace shadowing,³⁹ which is a qualitative research technique conducted on a small scale based on real-time observations. The team spent more than 80 hours with physicians, taking notes about barriers or behaviors associated with text entries to clinical notes. Their writing process was not interrupted during the shadowing phase. One of us (T.L.) collected the notes and then coded them using NVivo (Version 11).

System design

The hospital adopted an EHR platform, provided by Winning.⁴⁰ In March 2020, the vendor upgraded the platform from the Windows-based client to a thin client, implemented via the framework of Electron.⁴¹ The tech stack of the client was mainly based on JavaScript.

As most of desktop computers are on Windows operating system, Microsoft's Text Services Framework was a popular foundation for the architecture of IMEs.⁴² As standalone

tools, they are easy to install. But it was challenging for them to log entries within an EHR-driven context. With the help of "CompositionEvent,"⁴³ server-based logging was developed. The composition session for an IME entry was comprised of three phases, such as start, update, and end.⁴³ When a composition session starts, the event of "compositionstart" is fired. Each letter entered as query parameters would trigger a "compositionupdate." After a candidate is selected, "compositionend" is emitted. Then, client input behaviors were exposed to JavaScript. Some webhooks were developed for server-based collection. When the logging function was shifted from client to server, the context could be retrieved. Server-based logging was easy to configure, and it can be turned on or off centrally. The logging went alive after approval. After that, based on the architecture, the team worked on a prediction module. It was implemented as a visual intervention (Figure 4) at the front end. The detailed information was covered as a result in the section of "Suggestion Tool." It took six months for the development and testing. After that, the review board approved field-deployment.

Assessment methods

For evaluation, a parameter was designed to let each participant complete a testing using both approaches. After that, the parameter was automatically turned off for the participant. A training schedule was sent to physicians electronically. Nine hundred sixty-three physicians enrolled in an online workshop. The evaluation started after the training, by turning on the parameter for them. In the end, 952 physicians completed the trial within the four-week testing period.

For each participant, when an encounter was completed for note-taking using the intervention (Figure 4), the next encounter had to be completed with the traditional approach. With that said, each participant had to complete 100 notes for each approach during the evaluation. Entries collected with the traditional approach were classified as the control group, and the rest data was marked as the intervention group. All the participated physicians and patients were provided verbal informed consent.

For quantitative assessment, as shown in Table 1, Four metrics were designed for both of groups. For qualitative assessment of physician perceptions in terms of usefulness and trust, a questionnaire was designed, which contained an adapted version of the System Usability Scale⁴⁴ and a feature-level Likert-scale assessment. It was placed right after each participant completed testing. At last, 837 participants responded.

Results

Workspace findings

After the coding was completed, the team (P.W. and L.Y.) reviewed and some emergent codes were incorporated after

Table 1. Quantitative assessment metrics.

Metric	Description
Keystrokes per character (KPC)	For the study purpose, only clicks for Chinese characters were considered and the keystrokes used for selection (e.g. the Spacebar key) were excluded.
TLK rate	Based on keystroke intervals and selection keys, TLK was detected automatically, without an outlier.
Word per minute (WPM)	The number of Chinese characters entered within a minute.
Typo rate	The percentage of inputs completed with errors.

TLK, top-1 with less keystroke.

a discussion. Qualitative analysis of the shadowing notes resulted in 14 facilitators and 17 barriers. They are sorted and displayed in Table 2.

Word distribution

As shown in Figure 2, most of medical terms provided by Sougo IME dictionary have more than four characters. Based on logging, 81.7% of entries were completed in disyllable, 7.2% in monosyllable, 8.3% in three-syllable, and 2.7% in four-syllable. The shadowing results also suggested that physicians tended to split a longer input into multiple short semantic input units (SIUs). It suggested that the frequency was created on SIUs rather than the actual entity.

Context model

Although clinical notes promised expressivity for physicians,⁴⁵ the narrative often went through a semi-structured template with a couple of sections, such as subjective, objective, assessment, and plan.⁴⁶ \mathbb{G} was used to denote the context at the gross level. Some templates also contained subsections (e.g. physical exam).⁴⁷ Some of sections used a key phrase for enumerating a group of concepts.^{19,48} Thus, \mathbb{F} was used to denote the context at the fine level. Unlike English, Chinese does not use space for segmentation. With the IME-based text entry, a sentence was naturally segmented into multiple SIUs,⁴⁹ although truncation was sometimes used for postmodification. At the entity level, \mathbb{E} was used to denote the preceding component (Table 3). \mathbb{D} was used to denote the discipline for the template. A hierarchical semantic context was defined as $c = [\mathbb{D}, \mathbb{E}, \mathbb{F}, \mathbb{G}]$, as illustrated in Figure 3A.

Table 2. Facilitators and barriers.

Facilitators	Barriers
<ul style="list-style-type: none"> • Clinical notes follow a structured format • English words are mostly entered by turning off the IME • English words can be entered via the Enter from IMEs • Faster to use the Spacebar for selection • Key phrases are popularly used in sentences • Likely to try an abbreviated form of Pinyin when the intended word is a frequently used term • Likely to try an abbreviated form of Pinyin when the length of the intended word is more than 2 • Most of the input words are between 1 and 4 characters • Physicians are able to choose their favorite IMEs for text entry • Reduce the reaction time by typing a frequently word and then truncating one or two syllables • The abbreviated forms are learned through repeated practices, gradually increasing the length of input pinyin • The Backspace is touch-typing and used frequently for truncating • Uppercases are not treated as valid Pinyin • Use continuous typing to complete an input unit before stopping for disambiguation 	<ul style="list-style-type: none"> • An out-of-vocabulary word is limited to an individual computer after creation • Candidates are ranked by frequency, limited to an individual computer • Difficult for IMEs to access characters before the caret • Difficult for IMEs to navigate between the typed letters • Difficult for IMEs to retrieve the note structure • Easier to make errors on the front and the back-nasal sound, the retroflex, and the blade-alveolar • Learned frequency is lost after the computer is reinstalled • Learned frequency is not shared between different IMEs • Likely to cause typing errors on longer text entry • Not able to pursue a constant TLK when using the learned abbreviation for input • Patient privacy is at risk as the input is easy to track by the IME vendor • Slower to use the numeric keys for selection • Take extra time to fix typing errors • Take more time for reaction while navigating between letters that have been typed • The IME is a general tool for typing, not limited to clinical documentation • Word frequency is affected because of shared computers • Word frequency is affected because of varied documentation scenarios

TLK, top-1 with less keystroke.

For a given c , a word w was pending for documentation, physicians chose a specific form of Pinyin, denoted by p_s . After w was selected and inserted, the system retrieved all of the possible forms for p_s , denoted by $p_1, p_2, p_3, \dots, p_n$,

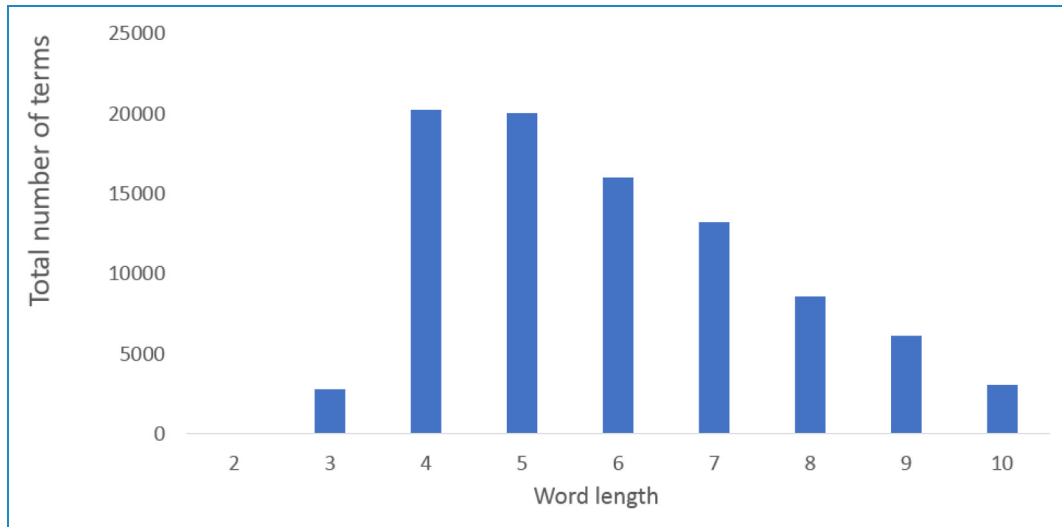


Figure 2. The distribution of medical terms provided by the Sogou IME. IME, input method engine.

Table 3. Normalizations of the preceding component.

Preceding component	Normalization
Digit	_D_
Non-Chinese letter	_F_
A normal SIU	SIU itself
Others (stop words, punctuations)	_O_

SIU, semantic input units.

where n was the density of the forms, being sorted from the shortest (i.e. acronym) to the longest (i.e. full pinyin). For each form p_i , the system created a tuple, denoted by $t_i = (c, w, p_i)$. The tuples were logged in EHRs for all physicians.

Suggestion generation

Workspace shadowing suggested that Pinyin length was gradually increased in the process of pursuing TLK. The index of the desired form was denoted by x . All of the suggestions were ranked by frequency. $T(p_x|c)$ was used to denote the top-1 candidate for Pinyin p_x under the context c . The exploration of TLK was expected to end with $w = T(p_x|c)$. To avoid confusions, sometimes, Chinese Pinyin needs explicit input (i.e. single quote mark) to separate syllables, even for acronym-based entries. For example, the acronym for the Chinese word of “postoperative” could be simply entered as “sh.” But as per IMEs the valid input must be completed as “s’h.”

To be friendly, the system accepted an acronym without a divider. $\mathcal{A}(p_x)$ was a function to validate whether p_x was a valid acronym, using a limited length $|p_x| \leq 4$. Given a context c , the search algorithm was illustrated in Figure 3B. When $|w| < |p_x|$, a valid acronym might be as same as a full Pinyin for a different word. When nothing was found in a given c , the system tried to reduce the dimension of c . The search continued until a result was found.

Suggestion tool

The visual intervention was shown in Figure 4. When physicians entered p_x , the system surfaced the search result using a web-based popup layer, just above the line of working sentence. Instead of hacking the inherent suggestions from IMEs, the system used the surfaced content as the top-1 suggestion. When the intended word was popped up, physicians could directly hit the Enter key. After that, the system inserted the intended Chinese word rather than the literal value of p_x . Normally, the below IME menu contained five candidates. With the popup layer, one more choice was added. Thus, there were two top positions used for fast choice selection, which was arranged from top to bottom. Physicians could gradually increase the length of p_x to zoom out on both of the positions in pursuit of TLK. When the popup layer was not helpful, physicians still were able to use the IME menu for word selection. The prediction had an average end-to-end latency of ≈ 0.5 milliseconds, which is instantaneous to physicians, responding far below the 100 ms threshold.⁵⁰

Basically, both IMEs and the system provided conversions for pure lowercase letters. As aforementioned, when

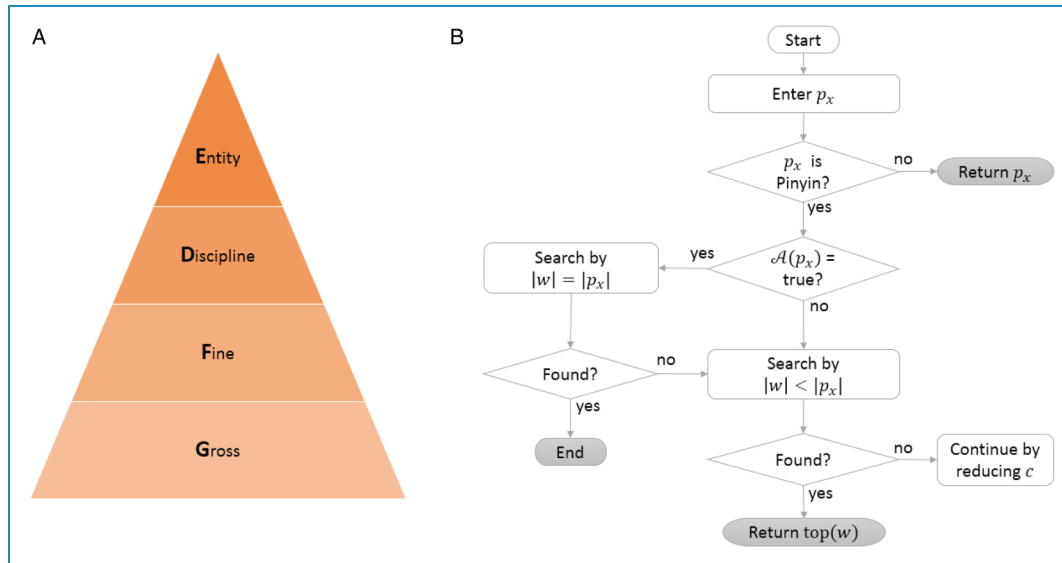


Figure 3. Proposed models. (A) Context model. (B) Suggestion generation.

the input sequence contained at least one uppercase, when the Enter was clicked, the system would skip suggesting and use the literal value of p_x for inserting. For a convertible p_x , if it was a valid English term defined in a dictionary, a message box popped up for confirmation. From there, the physician could choose whether it should be an English input or Chinese input.

Assessment results

The quantitative analysis results were shown in Table 4. In terms of keystrokes per character, compared to the control group, the intervention group received significant improvement ($p < .0001$). The average TLK rate was 0.62 for the control group. As observed in the shadowing phase, before the intervention, physicians tended to use full Pinyin and only used abbreviated Pinyin when they felt confident to bring the expected word to the top. The average rate was improved by 40.3% ($p < .0001$) in the intervention group. The overall word per minute was improved by 28.4%. After randomly selecting 100 records from each group, five typo errors were found in the control group and none was found in the intervention group. The qualitative analysis results were demonstrated in Figure 5. Overall, the team received positive feedback that the system was useful, though some physicians were not determined to use abbreviated Pinyin on infrequent terms.

We broke down the improvement in terms of KPC by note section. In descending order, they were review of systems (58.2%), diagnosis (58.1%), social history (58.0%), family history (57.8%), physical exam (57.7%), past medical history (57.3%), chief complaint (56.3%),

Table 4. Quantitative improvement.

Metric	Control	Intervention	Improvement	p-value
KPC, mean	2.9	1.5	48.3%	<.0001
TLK rate, mean	0.62	0.87	40.3%	<.0001
WPM, mean	55.2	70.9	28.4%	<.0001

TLK, top-1 with less keystroke; KPC, keystrokes per character; WPM, word per minute.

medications (52.6%), history of present illness (42.8%), and medical decision making (42.0%).

Discussion

Documentation burden is a leading factor for physician burnout, as too many keystrokes and time are spent on clinical text entry.^{3,4} For Chinese physicians, IMEs serve as an essential component for note-taking. Most of the IME-related time is spent on choice reaction and candidate selection.²⁶ Although abbreviated Pinyin saves keystrokes, full Pinyin is dominant as it helps reduce the time for choice reaction.³⁴ This study suggests an approach to alleviate the burden on both sides.

Unlike linguistic words, during the phase of input, words are segmented naturally by IMEs.⁴⁹ Based on the findings via system logs and workspace shadowing, this phenomenon is confirmed in the medical domain. Each IME word has 1–4 characters and has its own semantic meaning. Surprisingly, 81.7% of text entries are completed using



Figure 4. A screenshot of the suggestion tool.

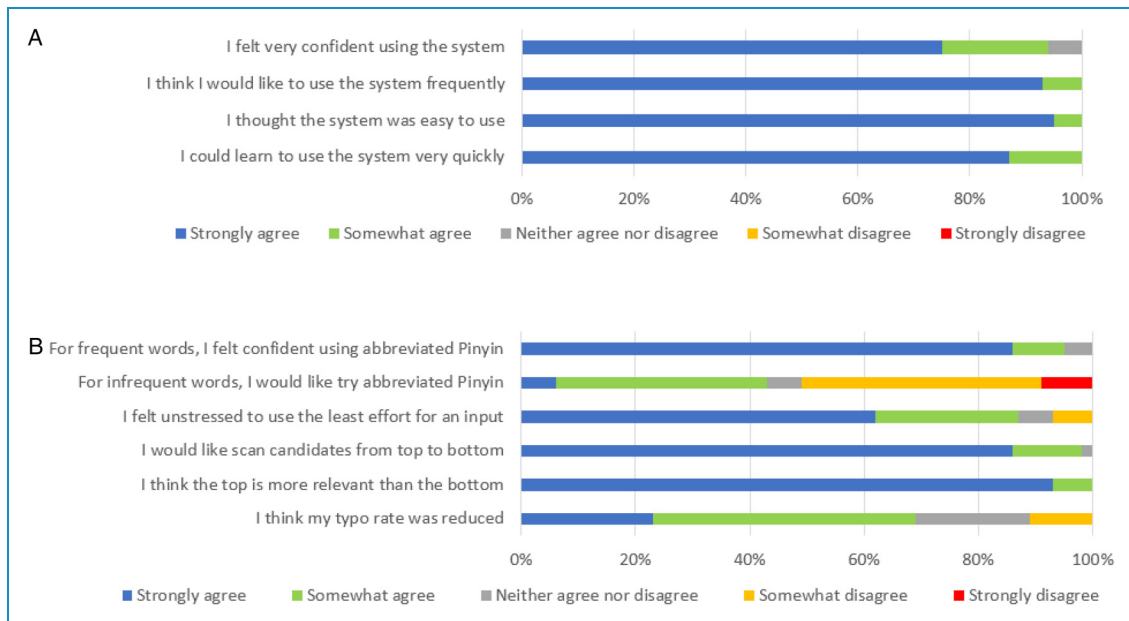


Figure 5. Usability evaluation results. (A) System usability scale assessment results. (B) Feature-level Likert-scale assessment results.

disyllables. The first reason is that 64% of the popular terms are two-character words.²⁷ The second reason is that disyllables are so elastic that they can be easily truncated to become a monosyllable or combined with the next entry to become a multisyllable. As observed from shadowing, the last reason might be related to the typo rate of Pinyin input.⁵¹ Typo rate grows with the length of the input sequence for Pinyin. When an error happens, it takes time and patience for finding out and correction. The splitting strategy forms the fundamental of the input language. A linguistic word with a lower term frequency is divided into multiple high frequently used IME words. This helps explaining why an extended medical dictionary would not boost the input performance as expected.³⁴

As short frequent IME words become the major part of typewriting, homophone disambiguation⁵² turns prominent. For both autocomplete and IMEs, as the triggering input sequence shrinks, the uncertainty of suggestions grows.

To improve top-1 accuracy while sticking to least effort of keystrokes, autocomplete tries to get more contextual information for help.^{19,20} On the other hand, autocomplete enhances accuracy based on a learned corpus coming from EHRs rather than an individual machine or physician. Fundamentally, the IME is a standalone tool, tethered to an individual machine. As a commercial product, it is not open-source to EHRs, even for the domain-specific IME.³³ As per the current IME architecture, EHRs are excluded for disambiguation. For security reasons, as a third-party tool, the IME has limited access to the context. As IMEs are used for daily input on all kinds of occasions, the number, sequence, and frequency distribution of the homophonic characters for each syllable are affected significantly,²⁶ especially on short frequent words. Consequently, 52% of entry time is reported to be an overhead,²⁶ used for choice reaction and locating selection keys. However, the disambiguation time is minimized

if the intended word is found at top-1, and selection is also touch-typing on the Spacebar key. It serves as the only entrance for TLK. Our study provides another entrance by unleashing the power of the Enter key. The dual entrance reserves two top positions for TLK from top to bottom. Besides, the popup layer is sensitive to the context. Compared to the control group, the intervention improves TLK rate by 40.3% ($p < .0001$). Meantime, keystrokes per character are reduced by 48.3% ($p < .0001$). Physicians also show great interest. The qualitative results suggest that a positive impact is happening to physician input behaviors. Being borrowed from autocomplete, the popup layer demonstrated the potential of bridging IMEs with EHRs. With that said, entries completed out of the IME menu are not tracked by the IME vendor. It helps protecting patient privacy.

Limitations

This study has some limitations. First, for simplicity and latency, rule-based prediction is used only by the proposed tool. Incorporating more contextual information (e.g. patient demographic and medical history) and deep learning could boost the top-1 accuracy. Second, the prediction is pragmatic based on collected system logs rather than a standard corpus. The results might vary in different settings. Third, the Spacebar is superior to the Enter in terms of touch-typing. However, being fixed by IMEs, the Spacebar cannot swap with the Enter, at least during text entry. But it suggests a direction for IMEs for future optimization. Fourth, the typo rate is evaluated qualitatively. A quantitative assessment is warranted in the future. Last, the project is being developed as a pragmatic improvement for a web-based EHR. The software is proprietary to the in-house platform.

Conclusions

Keystroke burden and frequent reaction time used for disambiguation challenge Chinese physicians in the process of routine text entry. Contextual autocomplete helps alleviating the burden in Western countries. For Chinese, IMEs are generally used for text entry, being separated from EHRs. This study suggests an approach to connect EHRs with IMEs. The bridge helps increasing the top-1 accuracy using keystrokes as few as possible. It is friendly to patient privacy and receives positive feedback from physicians.

Authors Note: Tao Li and Lei Yu are equally important as first authors for contribution.

Conflicting interests: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding: Shanghai Sixth People's Hospital, (grant number DY2019015)

Ethical approval: Not applicable.

Guarantor: P.W.

Contributorship: P.W., T.L., and L.Y. designed the study. P.W. developed the model and prepared figures. T.L. and L.Z. performed the evaluation and assembled the data. P.W., T.L., and L.Y. wrote the article. All authors contributed to the final approval of the article.

ORCID iD: Panzhang Wang  <https://orcid.org/0000-0002-2743-4904>

References

1. Friedberg MW, Chen PG, Van Busum KR, et al. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. *Rand Health Quarterly* 2014; 3: 1.
2. Ashton M. Getting rid of stupid stuff. *N Engl J Med* 2018; 379: 1789–1791.
3. Hill R, Sears LM and Melanson SW. 4000 clicks: a productivity analysis of electronic medical records in a community hospital ED. *Am J Emerg Med* 2013; 31: 1591–1594.
4. Verghese AC, Shah NH and Harrington RA. What this computer needs is a physician: humanism and artificial intelligence. *JAMA* 2018; 319: 19–20.
5. Chisholm CD, Weaver CS, Whenmouth LF, et al. A task analysis of emergency physician activities in academic and community settings. *Ann Emerg Med* 2011; 58: 117–122.
6. Khorana AA. Physician as typist. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology* 2010; 28: 3899–3900.
7. Reick-Mitrisin V, MacDonald M, Lin S, et al. Scribe impacts on US health care: benefits may go beyond cost efficiency. *J Allergy Clin Immunol* 2020; 145: 479–480.
8. Coiera E, Kocaballi B, Halamka J, et al. The digital scribe. *NPJ Digit Med* 2018; 1: 58.
9. Quiroz JC, Laranjo L, Kocaballi AB, et al. Challenges of developing a digital scribe to reduce clinical documentation burden. *NPJ Digit Med* 2019; 2: 114.
10. Wang J, Yang J, Zhang H, et al. Phenopad: building AI enabled note-taking interfaces for patient encounters. *NPJ Digital Medicine* 2022; 5: 12.
11. Blackley SV, Schubert VD, Goss FR, et al. Physician use of speech recognition versus typing in clinical documentation: a controlled observational study. *Int J Med Inf* 2020; 141: 104178.
12. Smith CA, Hetzel SJ, Dalrymple PW, et al. Beyond readability: investigating coherence of clinical text for consumers. *J Med Internet Res* 2011; 13.
13. Wang CJ. Medical documentation in the electronic era. *JAMA* 2012; 308: 2091–2092.
14. Xu H, Stetson PD and Friedman C. A study of abbreviations in clinical notes. *AMIA Annual Symposium Proceedings AMIA Symposium* 2007: 821–825.

15. Stevenson M, Guo Y, Alamri A, et al. Disambiguation of Biomedical Abbreviations. Paper presented at: BioNLP@HLT-NAACL2009.
16. Skreta M, Arbabi A, Wang J, et al. Automatically disambiguating medical acronyms with ontology-aware deep learning. *Nat Commun* 2021; 12.
17. Gerard M, Chimowitz H, Fossa A, et al. The importance of visit notes on patient portals for engaging less educated or nonwhite patients: survey study. *J Med Internet Res* 2018; 20.
18. Koch-Weser S, DeJong W and Rudd RE. Medical word use in clinical encounters. *Health Expect* 2009; 12.
19. Gopinath D, Agrawal M, Murray LS, et al. Fast, Structured Clinical Documentation via Contextual Autocomplete. Paper presented at: MLHC2020.
20. Greenbaum NR, Jernite Y, Halpern Y, et al. Improving documentation of presenting problems in the emergency department using a domain-specific ontology and machine learning-driven user interfaces. *Int J Med Inform* 2019; 132: 103981.
21. Murray LS, Gopinath D, Agrawal M, et al. MedKnows: Unified Documentation and Information Retrieval for Electronic Health Records. ArXiv 2021: abs/2109.11451.
22. Yamaguchi M and Logan GD. Pushing typists back on the learning curve: memory chunking in the hierarchical control of skilled typewriting. *Journal of Experimental Psychology Learning, Memory, and Cognition* 2016; 42: 1919–1936.
23. MacKenzie IS. Fitts' law as a research and design tool in human-computer interaction. *Hum Comput Interact* 1992; 7: 91–139.
24. Zheng Y, Li C and Sun M. CHIME: an Efficient Error-Tolerant Chinese Pinyin Input Method. Paper presented at: IJCAI2011.
25. Siok WT and Liu CY. Differential impacts of different keyboard inputting methods on reading and writing skills. *Sci Rep* 2018; 8.
26. Wang J, Zhai S and Su H. Chinese input with keyboard and eye-tracking: an anatomical study. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* 2001; 3: 349–356.
27. *Lexicon of Contemporary Chinese, Third version, ISBN 7-100-01777-7/H.519*. Beijing: Shangwu Publishing House, 1996.
28. Suzuki H and Gao J. A Unified Approach to Transliteration-based Text Input with Online Spelling Correction. Paper presented at: EMNLP2012.
29. Liu Y and Wang Q. Chinese pinyin phrasal input on mobile phone: usability and developing trends. Paper presented at: Mobility '072007.
30. *Sogou Pinyin*. <https://pinyin.sogou.com/>. Accessed.
31. *Baidu Input Method Software*. <https://shurufa.baidu.com/>. Accessed.
32. *Google Pinyin*. <https://www.google.com/intl/zh-CN/inputtools/>. Accessed.
33. *Sogou Pinyin Doctor Edition*. <https://pinyin.sogou.com/yisheng>. Accessed.
34. Yang F, Guo H and Li J. TestIME: an application for evaluating the efficiency of Chinese input method engines in electronic medical record entry task. *BMC Med Inform Decis Mak* 2019; 19.
35. Huang Y, Li Z, Zhang Z, et al. Moon IME: Neural-based Chinese Pinyin Aided Input Method with Customizable Association. Paper presented at: ACL2018.
36. Zhang X, Wei C and Hai Z. Tracing a Loose Wordhood for Chinese Input Method Engine. ArXiv. doi:2017;abs/1712.04158.
37. Huang Y and Hai Z. Chinese Pinyin Aided IME, Input What You Have Not Keystroked Yet. Paper presented at: EMNLP2018.
38. Huang Y, Zhang Z and Zhao H. Neural-based Pinyin-to-Character Conversion with Adaptive Vocabulary. ArXiv. 2018;abs/1811.04352.
39. Kitsis E. Shining a light on shadowing. *JAMA* 2011; 305: 1029–1030.
40. *Winning EHR*. <https://www.winning.com.cn/>. Accessed.
41. Build cross-platform desktop apps with JavaScript, HTML, and CSS. <https://www.electronjs.org/>. Accessed.
42. *Microsoft's Text Services Framework (TSF)*. <https://docs.microsoft.com/en-us/windows/desktop/tsf/text-services-framework>. Accessed.
43. *CompositionEvent*. <https://developer.mozilla.org/en-US/docs/Web/API/CompositionEvent>. Accessed.
44. Brooke JB. SUS: A 'Quick and Dirty' Usability Scale. 1996.
45. Rosenbloom ST, Denny JC, Xu H, et al. Data from clinical notes: a perspective on the tension between structure and flexible documentation. *J Am Med Inform Assoc* 2011; 18: 181–186.
46. Cameron S and Turtle Song I. Learning to write case notes using the SOAP format. *J Couns Dev* 2002; 80: 286–292.
47. Haug PJ, Wu X, Ferraro JP, et al. Developing a section labeler for clinical documents. *AMIA Annu Symp Proc* 2014; 2014: 636–644.
48. Colicchio TK, Dissanayake PI and Cimino JJ. The anatomy of clinical documentation: an assessment and classification of narrative note sections format and content. *AMIA Annu Symp Proc* 2020; 2020: 319–328.
49. Zhang D, Mao Y, Liu YP, et al. The Discovery of Natural Typing Annotations: User-produced Potential Chinese Word Delimiters. Paper presented at: ACL2015.
50. Nielsen J. Usability engineering. Paper presented at: the Computer Science and Engineering Handbook1997.
51. Zheng Y, Xie L, Liu Z, et al. Why Press Backspace? Understanding User Input Behaviors in Chinese Pinyin Input Method. Paper presented at: ACL2011.
52. Huff C and Payne E. Disambiguation of Tonally Unspecified Mandarin Syllables. 2012.