A Cost–Benefit Study of Text Entry Suggestion Interaction

Philip Quinn[1,2]   **Shumin Zhai**[1]

[1]Google Inc.
Mountain View, CA, USA

[2]University of Canterbury
Christchurch, New Zealand

philip@quinn.gen.nz, zhai@google.com

ABSTRACT

Mobile keyboards often present error corrections and word completions (*suggestions*) as candidates for anticipated user input. However, these suggestions are not cognitively free: they require users to attend, evaluate, and act upon them. To understand this trade-off between suggestion savings and interaction costs, we conducted a text transcription experiment that controlled *interface assertiveness*: the tendency for an interface to present itself. Suggestions were either always present (*extraverted*), never present (*introverted*), or gated by a probability threshold (*ambiverted*). Results showed that although increasing the assertiveness of suggestions reduced the number of keyboard actions to enter text and was subjectively preferred, the costs of attending to and using the suggestions impaired average time performance.

Author Keywords
Interface assertiveness; predictive interfaces; mobile text entry.

ACM Classification
H.1.2 [**User/Machine Systems**] Human factors.

INTRODUCTION

A technique frequently employed to improve user performance is to observe a user's interactions, and detect potential errors or predict intended interactions from their behaviour. Corrections for these errors and candidates for these intentions (collectively *suggestions*) can be offered to the user as shortcuts that optimise away interactions the system can anticipate. However, although these suggestions are frequently employed, their *costs* and *benefits* are often not systematically studied [cf. 25]. That is, are the potential benefits of a suggestion worth the costs of interrupting the user, the time and effort for them to process it, decide if it is accurate, and ultimately accept or reject it?

A particularly intense use of suggestion interfaces is with mobile touch-keyboard interaction, where there are many opportunities to adapt the keyboard interface to enhance user performance [e.g. 5, 22, 32, 33]. Text entry suggestions attempt to relieve the burden of precise and tedious character

input by offering predictions of the user's intended words or corrections for errors in words that have already been entered. The goal is to reduce the number of keystrokes or corrective actions required to compose text. However, their use is not perceptually or cognitively free: users must identify and attend to the display of the suggestions, evaluate their correctness, and possibly accept or reject them [13, 15].

A key consideration for these interfaces is ultimately *when* to present suggestions to the user. Many systems maintain an internal confidence value associated with each prediction (determined from a language model or dictionary) and only present those above a certain threshold. However, an aspect that is often overlooked is the overall *pay-off* or *utility* to the user – that is, does the user benefit over continuing to type the word or correcting the error manually? Benefits include the time from tapping on the requisite keys, but must be reconciled with the costs of attending and interacting with the interface.

In this paper, we term the tendency for an interface to present suggestions to a user its *assertiveness*, and examine its influence on user performance and satisfaction. Using such a control may improve user experience by reducing the amount of distraction and maximising the benefit of the suggestions. We examine the costs and benefits of text-entry suggestion interaction in an experiment that examines objective text-entry performance and subjective preference at different levels of assertiveness for word-completion suggestions. We find that increasing assertiveness decreases time performance but increases input economy and user preference; we discuss the implications for suggestion interface design.

BACKGROUND
Mobile soft keyboards are rendered and interacted with directly on a display, using either a finger on a touch-sensitive surface or a stylus with a digitiser [19, 21]. However, compared to their physical counterparts, the lack of strong tactile feedback limits users' ability to touch type [2, 10, 23, 28, 32].

Suggestion Systems
Prediction and error correction systems aim to reduce the number of actions required from a user to enter their desired text [4, 7, 8, 24, 31, 32]. Augmentative and Alternative Communication (AAC) research has investigated prediction systems to improve input speed for users with physical impairments, but has found that the benefits of suggestions are not always clear [8, 31]. Koester and Levine [13–15] suggested that the cognitive and motor costs of using suggestion interfaces may outweigh their benefits, and their evaluations of several AAC

systems found either marginal performance gains for novice users, or performance impairments for experienced users.

Current commercial practice makes use of suggestion interfaces by default. Google's Android platform (version 5 'Lollipop') persistently presents up to three suggestions above the keyboard. In contrast, Apple's iOS platform (version 9) can display one suggestion below the text insertion cursor, or optionally, three in a bar above the keyboard. The suggestion below the cursor is transient, and is accepted by tapping on the spacebar (tapping on the suggestion will reject it).

Models of Interaction

Interaction with soft keyboards has been successfully modelled using elemental cognitive models (such as GOMS/KLM [3]) in a chain of key-tapping actions [e.g. 18, 29, 30]. These models have been extended to cover the actions involved in using suggestion interfaces [e.g. 11, 14, 17]. However, such models predict performance for expert users and focus on the mechanical aspects of interaction – with limited scope for the cognitive processes or novice users [cf. 1]. Models of attention [e.g. 12] capture issues of interruption and distraction, but have focussed on distractions from unrelated tasks (e.g. email notifications) rather than potentially helpful suggestions for the current task. However, recent work has begun to explore the interaction between a system's attempts at assistance and subjective user preferences [25].

INTERFACE ASSERTIVENESS

Suggestions are a problem for interaction design because of their probabilisitic nature: they are a *guess* about a user's intentions. Incorrect guesses will waste a user's time and lead them to spurn the interface (disabling or habitually ignoring its suggestions). To create a productive experience, it is therefore desirable to be maximally beneficial when suggestions are correct, with minimal costs when they are incorrect.

To use a suggestion, a user must: (a) recognise the presence of the suggestion(s), (b) evaluate the options presented, and (c) optionally take action upon them [13]. As the cost of these actions may be high and provide little benefit (distracting the user for potentially incorrect suggestions), determining precisely *when* to show suggestions should be assiduously considered. For example, suggestions may only be shown if they pass some confidence threshold (typically, according to an *a posteriori* estimation from a language model). In these cases, the user will only need to attend to the suggestions that system decides they are worth the cost. Similarly, a system may mediate suggestions with an H-metaphor control system that is responsive to changes in user behaviour [6].

A method for assessing these costs is the *utility* of the suggestions to be presented: a consideration of system's confidence that the suggestion is correct, weighted with the predicted time savings offered to the user (e.g. time saved tapping keys or editing the text) against the costs of attending to the suggestion. For example, completing the last character of a word does not offer much utility to a user as they may be able to hit the key for that character faster than they could attend to the suggestion interface, evaluate the suggestion, and confirm it.

Suggestion Presentation

The quality of suggestions depends on the system's method for generating them – such as a language model, adaptation from a user's behaviour [7], or characteristics of their input [6, 32]. However, while the quality of suggestions is a key factor in user performance, so too is how the system communicates them and how users interact with them.

Location. Where does the interface display suggestions? The user's awareness and attention to certain parts of the screen may influence whether they notice the suggestions, and alter the interaction requirements for using with them.

Presence. How many suggestions are shown at once? Each suggestion presented increases cognitive costs, as the user must evaluate its correctness. Showing only the top suggestion reduces this, but at the risk of ignoring good secondary options.

Default Behaviour. Are suggestions with a high probability of being correct automatically accepted? This eliminates the cost of explicitly accepting a suggestion, but at the cost of requiring the user to explicitly reject incorrect suggestions or manually correct accidental acceptances.

Stability. Are suggestions continually changing, or are they stable and minimise visual disruptions? Similarly, does the system learn from the user's use of them to help develop automaticity? High stability promotes rapid use of the interface and reduces cognitive costs, but risks hiding good suggestions.

These factors may in-turn influence the underlying model for choosing suggestions. For example, if only a single suggestion is presented, the system may want to ensure that there is a large confidence gap between the first and remaining options. Similarly, if the user has to explicitly reject poor suggestions, then the confidence threshold may be higher.

EXPERIMENT

We conducted an experiment to better understand the trade-offs involved in text entry suggestions in terms of user performance and subjective preference. Subjects copied simple phrases, with word completions shown at opportune times. Completions were ranked by a probability function that emulated current practice, and were presented to subjects in one of three assertiveness conditions: always (*extraverted*), never (*introverted*), and gated by a threshold (*ambiverted*).

Subjects & Apparatus

Seventeen (five female) volunteer subjects participated in the experiment; all had experience typing on mobile keyboards (from various vendors) and received a gift card for their time.

The experiment was run on an Apple iPod Touch (A1367) under iOS 5.1.1 (640×960 px screen at 326 ppi). All visual ornaments were hidden and default text editing features suppressed. Feedback was identical to the English iOS keyboard, but with the modifier keys removed (Figure 1). Letter keys had a visual size of 52×76 px, but their targetable area was expanded to consume the surrounding space ($\sim 65 \times 108$ px).

Stimuli & Tasks

Subjects copied phrases using the keyboard, with up to three suggestions occasionally appearing directly above it (Figure 1).
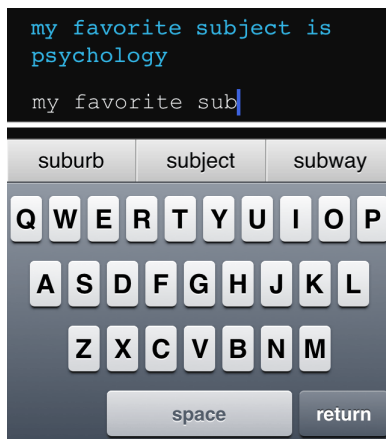
Figure 1. The experimental interface.

Any combination of correct letter keys and helpful suggestions could be used to complete the phrase.

Input. Suggestion interfaces are dependent upon a user's input accuracy to offer predictions: it difficult to generate good predictions for error-laden input, while error-free input eliminates any need for them. The input accuracy of subjects is not of interest here – only their interaction with the suggestion interface is. Therefore, we artificially constrained input to be error-free and only examined word-completion suggestions (not error corrections). This was accomplished by only accepting taps on the correct key or suggestion (ignoring all other taps with no feedback). Combined with the probability scoring (described below), this controlled the occurrence and content of suggestions.

Phrases & Suggestions. Phrases were randomly selected from a set of 500 [20] – modified for consistent spelling in all-lowercase, and without punctuation. Suggestions were generated from the open-source Google Android dictionary.[1]

Probability Scoring. Frequency information from the dictionary determined an estimate of suggestion probability: $(f/255) \cdot (1 - (c/s))$, where f is the frequency of the word (in the range $[0, 255]$), s is its length, and c is the number of letters completed by it.[2] The estimate was normalised by the sum of all candidate estimates. With the given dictionary, this has a theoretical average saving of 46% of the letters in each word (i.e. the correct suggestion appears in the top three after the user enters approximately half of the letters of a word).

Suggestions. After each key tap, the interface scored suggestions matching the entered word prefix. In the *extraverted* condition, or if any of the scores were above a threshold of 0.1 in the *ambiverted* condition, the top three suggestions (sorted by score) were displayed. If all words were below the threshold, any currently displayed suggestions remained. The suggestion bar was cleared between words.

Suggestions were selected by tapping on them, whereupon the characters it completed were inserted into the response.

---

[1]`https://goo.gl/wN4vHn`
[2]A simplified version of the scoring method used by Google Android:
`https://goo.gl/vk4XNh`, lines 1098–1147.

Design & Procedure
A within-subjects design was used for the factor *assertiveness* ∈ *{extraverted, ambiverted, introverted}*, with the order of presentation counterbalanced. Subjects were instructed to enter the text however they felt was fast and comfortable, and should not feel compelled to use the suggestions.

Subjects first completed five practice phrases in an *extraverted* condition. For each condition, they completed 18 randomly selected phrases (without replacement), with the first three discarded – followed by a NASA-TLX questionnaire [9].

RESULTS
There are many metrics that can assess typing performance [19, 21]. We analyse task completion time measured in characters per second (CPS) for each phrase, but are also interested in the number of taps (on keys and suggestions) required per character of input (TPC), and the presentation/usage of the suggestion interface.

Performance
Figure 2a shows the mean CPS, and a one-way ANOVA for *assertiveness* revealed a significant effect ($F_{2,32} = 12.16$, $p < .001$, $\eta_G^2 = 0.07$), with *introverted* (mean 3.09 CPS, 95% CI ±0.12) outperforming both *ambiverted* (2.81 ± 0.09) and *extraverted* (2.66 ± 0.09). For the number of taps required per character (TPC; Figure 2b), a Friedman test revealed a significant effect ($\chi_2^2 = 34$, $p < .001$), with *extraverted* requiring significantly fewer (mean .93 TPC, 95% CI [.92, .93]) taps than either *ambiverted* (.96 [.95, .96]) or *introverted* (1 [1, 1]); Bonferroni-corrected Wilcoxon signed-rank tests revealed significant differences ($p \leq .003$) between all pairs.

Suggestions
While suggestions in the *extraverted* condition were updated at every opportunity (after every letter key tap or suggestion selection), they were updated after only 26.72% of such opportunities in the *ambiverted* condition (based on the threshold). Figure 2c shows the proportion of those opportunities where suggestions were shown, those where helpful suggestions were shown,[3] and those where subjects utilised them.

The tap savings that subjects could have obtained if they selected the correct suggestion at the earliest opportunity and the savings actually realised are shown in Figure 2d. Wilcoxon signed-rank tests between the *extraverted* and *ambiverted* conditions revealed significant differences for both the savings offered and realised (both $W = 153$, $p < .001$). In both cases, *extraverted* provided higher potential (17.59% vs. 9.44%) and realised (9.44% vs. 4.66%) tap savings than *ambiverted*.

Subjective Responses
Responses from the NASA-TLX surveys are shown in Table 1. Friedman tests revealed significant differences for *physical demand* ($\chi_2^2 = 14.31$) and *effort* ($\chi_2^2 = 16.04$, both $p \leq .001$). Bonferroni-corrected Wilcoxon signed-rank tests found those ratings for *introverted* to be significantly higher.

DISCUSSION
Our results found that an *ambiverted* interface – guarding suggestions by a threshold – was faster, on average, than an

---

[3]Less than the theoretical 46% due to the relatively short words in the experimental phrases and consideration of the tap to select a suggestion.
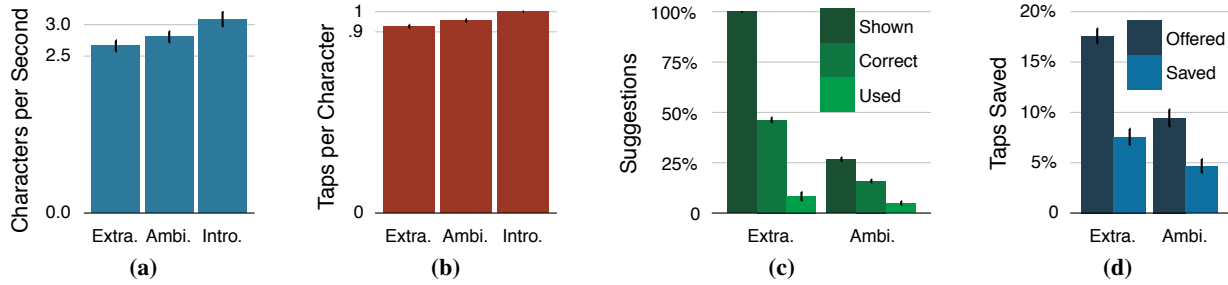
Figure 2. Mean performance metrics (±95% CIs) for each condition: (a–b) entry speed; (c) the proportion of suggestions bars shown, that contained a helpful suggestion, and that were actually used; and (d) the potential and obtained tap savings.

	Mental	**Phys.**	Temp.	Perf.	**Effort**	Frust.
Extra.	2.41	2.24	1.94	2.65	2.24	1.94
	(0.94)	(0.83)	(0.56)	(1.06)	(0.56)	(0.75)
Ambi.	2.00	2.18	1.00	3.12	2.53	2.29
	(1.06)	(0.88)	(0.79)	(1.17)	(1.12)	(0.99)
Intro.	2.12	3.41	2.47	3.06	3.59	2.59
	(1.27)	(1.12)	(1.28)	(1.03)	(1.00)	(1.33)

Table 1. Mean (and std. dev.) NASA-TLX ratings (scale: 1–5).

extraverted interface that always made suggestions visible; and an *introverted* baseline that never made any suggestions was faster still. The performance loss with suggestions was in spite of the reduced average number of taps required to enter text. Nevertheless, suggestions were subjectively considered less physically demanding and effortful.

This result follows that of prior work in a different context (motor-control-impaired users on physical keyboards [14, 15]), which found it difficult for suggestions to gain a time advantage over the perceptual and cognitive costs they introduced. Our results are based on a modern touchscreen keyboard, and the constraint of perfect input from subjects arguably gives a greater time advantage to using suggestions. The maximum theoretical letter savings of 46% is similar to the current limit of production interfaces [7], but it is likely that developments in prediction systems will change this cost-benefit balance.

Although the time cost of attending, deciding on, and selecting suggestions lowered average speed, subjective comments following the experiment (in addition to TLX responses) indicated a strong dislike for the *introverted* condition: that it felt more demanding and time consuming. This suggests that the user experience benefits of suggestions lie beyond average time performance (e.g. spelling assistance or learned automaticity), and may be of greater psychological value than their objective costs imply. Therefore, we cannot conclude that that less assertiveness necessarily results in a better user experience [25].

Our results are also limited to word completion suggestions. *N*-gram word prediction and correction of input errors may provide greater savings at the same or similar costs of attending, deciding, and selecting them. Such suggestions may take

advantage of natural pauses in user input, or the high costs of editing already entered text on mobile interfaces. The experimental task was also a copying task, where subjects could always see the exemplar text to be entered – and were not encumbered by decisions about how to *compose* the text, or uncertainty about spelling or grammar [16]. When a user is unsure about the correct spelling of a word, a correctly spelt suggestion may be substantially faster than trying out spellings and editing the text when it is discovered to be incorrect.

Suggestion interfaces may also have an *overhead cost* in addition to the costs of using them. That is, the anticipation of making a decision, may introduce a persistent cost to the use of the interface. Such overhead costs are likely to be interface and user-dependent, but are critical to understand if systems are to be designed to overcome them [e.g. 26, 27]. Conversely, common suggestions for habitual errors may be learned and become an automatic part of a user's behaviour and removing them may impair performance and overall experience. Teasing apart these factors will require detailed and precise experimentation.

Our results have shown that suggestion interfaces face significant challenges in enabling users to improve their input performance. However, the costs associated with suggestions deserve more attention, as do more principled approaches for managing assertiveness (e.g. with models of economic risk and utility). Researchers and practitioners developing systems with such trade-offs should be cognisant of these costs, but also that benefits may be realised in other types of assistance.

CONCLUSIONS & FUTURE WORK

We have presented an empirical analysis of interaction with text input suggestions, focussing on how eager interfaces should be to show word completions (their *assertiveness*). A moderately assertive (*ambiverted*) interface that guarded completions with a confidence threshold afforded faster performance than a more assertive (*extraverted*) interface that always presented the best three. While both types of interface were, on average, slower than never presenting completions at all (*introverted*), the suggestions were utilised to save taps and were subjectively preferred – indicating benefits not measured by time. Future work will examine broader text entry tasks (e.g. composition and error correction), and model this trade-off by examining the *utility* of text entry suggestions.

REFERENCES

1. Ahmed Sabbir Arif and Wolfgang Stuerzlinger. 2010. Predicting the cost of error correction in character-based text entry technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, 5–14. DOI: http://dx.doi.org/10.1145/1753326.1753329

2. Shiri Azenkot and Shumin Zhai. 2012. Touch behavior with different postures on soft smartphone keyboards. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services (MobileHCI '12)*. ACM, New York, NY, 251–260. DOI: http://dx.doi.org/10.1145/2371574.2371612

3. Stuart K. Card, Thomas P. Moran, and Allen Newell. 1983. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, London.

4. Mark Dunlop and John Levine. 2012. Multidimensional pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, 2669–2678. DOI: http://dx.doi.org/10.1145/2207676.2208659

5. Leah Findlater and Jacob Wobbrock. 2012. Personalized input: improving ten-finger touchscreen typing through automatic adaptation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, 815–824. DOI: http://dx.doi.org/10.1145/2207676.2208520

6. Frank O. Flemisch, Catherine A. Adams, Sheila R. Conway, Ken H. Goodrich, Michael T. Palmer, and Paul C. Schutte. 2003. *The H-Metaphor as a Guideline for Vehicle Automation and Interaction*. Technical Report NASA/TM-2003-212672. NASA Langley Research Center, Hampton, VA. http://ntrs.nasa.gov/search.jsp?R=20040031835

7. Andrew Fowler, Kurt Partridge, Ciprian Chelba, Xiaojun Bi, Tom Ouyang, and Shumin Zhai. 2015. Effects of Language Modeling and Its Personalization on Touchscreen Typing Performance. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, 649–658. DOI: http://dx.doi.org/10.1145/2702123.2702503

8. Nestor Garay-Vitoria and Julio Abascal. 2006. Text prediction systems: a survey. *Universal Access in the Information Society* 4 (2006), 188–203. DOI: http://dx.doi.org/10.1007/s10209-005-0005-9

9. Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. Elsevier, North Holland, 139–183. DOI: http://dx.doi.org/10.1016/S0166-4115(08)62386-9

10. Niels Henze, Enrico Rukzio, and Susanne Boll. 2012. Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, 2659–2668. DOI: http://dx.doi.org/10.1145/2207676.2208658

11. Heidi M. Horstmann and Simon P. Levine. 1990. Modeling of user performance with computer access and augmentative communication systems for handicapped people. *Augmentative and Alternative Communication* 6, 4 (1990), 231–241. DOI: http://dx.doi.org/10.1080/07434619012331275514

12. Eric Horvitz, Carl Kadie, Tim Paek, and David Hovel. 2003. Models of attention in computing and communication: from principles to applications. *Commun. ACM* 46, 3 (March 2003), 52–59. DOI: http://dx.doi.org/10.1145/636772.636798

13. Heidi Horstmann Koester and Simon P. Levine. 1993. A model of performance cost versus benefit for augmentative communication systems. In *Proceedings of the 15th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 1303–1304. DOI: http://dx.doi.org/10.1109/IEMBS.1993.979147

14. Heidi Horstmann Koester and Simon P. Levine. 1994. Modeling the Speed of Text Entry with a Word Prediction Interface. *IEEE Transactions on Rehabilitation Engineering* 2, 3 (1994), 177–187. DOI: http://dx.doi.org/10.1109/86.331567

15. Heidi Horstmann Koester and Simon P. Levine. 1996. Effect of a word prediction feature on user performance. *Augmentative and Alternative Communication* 12, 3 (1996), 155–168. DOI: http://dx.doi.org/10.1080/07434619612331277608

16. Per Ola Kristensson and Keith Vertanen. 2012. Performance comparisons of phrase sets and presentation styles for text entry evaluations. In *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces (IUI '12)*. ACM, New York, NY, 29–32. DOI: http://dx.doi.org/10.1145/2166966.2166972

17. Ying Liu and Kari Jouko Räihä. 2010. Predicting Chinese text entry speeds on mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, 2183–2192. DOI: http://dx.doi.org/10.1145/1753326.1753657

18. I. Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. 2001. Accuracy measures for evaluating computer pointing devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '01)*. ACM, New York, NY, 9–16. DOI: http://dx.doi.org/10.1145/365024.365028

19. I. Scott MacKenzie and R. William Soukoreff. 2002. Text Entry for Mobile Computing: Models and Methods, Theory and Practice. *Human-Computer Interaction* 17, 2–3 (2002), 147–198. DOI: http://dx.doi.org/10.1080/07370024.2002.9667313

20. I. Scott MacKenzie and R. William Soukoreff. 2003. Phrase sets for evaluating text entry techniques. In *CHI '03 Extended Abstracts on Human Factors in*

Computing Systems (CHI EA '03). ACM, New York, NY, 754–755. DOI:`http://dx.doi.org/10.1145/765891.765971`

21. I. Scott MacKenzie and Kumiko Tanaka-Ishii (Eds.). 2007. *Text Entry Systems: Mobility, Accessibility, Universality*. Morgan Kaufmann, San Francisco, CA.

22. Toshiyuki Masui. 1998. An efficient text input method for pen-based computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., New York, NY, 328–335. DOI:`http://dx.doi.org/10.1145/274644.274690`

23. Hugo Nicolau and Joaquim Jorge. 2012. Touch typing using thumbs: understanding the effect of mobility and hand posture. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, 2683–2686. DOI:`http://dx.doi.org/10.1145/2207676.2208661`

24. Antti Oulasvirta, Anna Reichel, Wenbin Li, Yan Zhang, Myroslav Bachynskyi, Keith Vertanen, and Per Ola Kristensson. 2013. Improving two-thumb text entry on touchscreen devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, 2765–2774. DOI:`http://dx.doi.org/10.1145/2470654.2481383`

25. Philip Quinn and Andy Cockburn. 2016. When Bad Feels Good: Assistance Failures and Interface Preferences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY. DOI:`http://dx.doi.org/10.1145/2858036.2858074`

26. Philip Quinn, Andy Cockburn, and Jérôme Delamarche. 2013. Examining the costs of multiple trajectory pointing techniques. *International Journal of Human-Computer Studies* 71, 4 (2013), 492–509. DOI:`http://dx.doi.org/10.1016/j.ijhcs.2012.12.001`

27. Philip Quinn, Andy Cockburn, Kari-Jouko Räihä, and Jérôme Delamarche. 2011. On the costs of multiple trajectory pointing methods. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, 859–862. DOI:`http://dx.doi.org/10.1145/1978942.1979067`

28. Andrew Sears. 1991. Improving Touchscreen Keyboards: Design issues and a comparison with other devices. *Interacting with Computers* 3, 3 (1991), 253–269. DOI:`http://dx.doi.org/10.1016/0953-5438(91)90016-U`

29. Andrew Sears, Julie A. Jacko, Josey Chu, and Francisco Moro. 2001. The role of visual search in the design of effective soft keyboards. *Behaviour & Information Technology* 20, 3 (2001), 159–166. DOI:`http://dx.doi.org/10.1080/01449290110049790`

30. R. William Soukoreff and I. Scott MacKenzie. 1995. Theoretical upper and lower bounds on typing speed using a stylus and a soft keyboard. *Behaviour & Information Technology* 14, 6 (1995), 370–379. DOI:`http://dx.doi.org/10.1080/01449299508914656`

31. Keith Trnka, John McCaw, Debra Yarrington, Kathleen F. McCoy, and Christopher Pennington. 2009. User Interaction with Word Prediction: The Effects of Prediction Quality. *ACM Transactions on Accessible Computing* 1, 3, Article 17 (Feb. 2009), 34 pages. DOI:`http://dx.doi.org/10.1145/1497302.1497307`

32. Daryl Weir, Henning Pohl, Simon Rogers, Keith Vertanen, and Per Ola Kristensson. 2014. Uncertain Text Entry on Mobile Devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, 2307–2316. DOI:`http://dx.doi.org/10.1145/2556288.2557412`

33. Shumin Zhai, Michael Hunter, and Barton A. Smith. 2002. Performance Optimization of Virtual Keyboards. *Human-Computer Interaction* 17, 2–3 (2002), 229–269. DOI:`http://dx.doi.org/10.1080/07370024.2002.9667315`