Barriers to Adoption of Dictionary-Based Text-Entry Methods: A Field Study

Howard Gutowitz
Eatoni Ergonomics, Inc.
42 W. 24th Street
New York, New York 10010

hag@eatoni.com

Abstract

MacKenzie et al. (2001) published user-study and simulation results which suggest that text-entry speed on cell phone keypads backed by predictive dictionary-based software is highly dependent on the fraction of words the user needs which are present in the dictionary. They found that when the fraction of non-dictionary words is greater than approximately 15% percent, dictionary-based methods are slower than multi-tap.

To assess the relevance of these laboratory findings to user experience in practice, we conducted a field study of the use of these dictionary-based methods. The field study comprised 230 interviews with people who had cell phones and who had some experience with dictionary-based text-entry methods. Participants were asked questions concerning their use of messaging and of a dictionary-based method, if any. Non-users most frequently cite difficulty of use or dictionary deficiencies as their reason for not using a dictionary-based method, and dictionary deficiencies are also often cited by users. We supply evidence that dictionary deficiencies contribute to making the systems hard to learn as well as hard to use. A simulation was performed to account for this latter result. In this survey, somewhat under half of those exposed to predictive text use it. This is to be compared to the results of Döring (2002) who reports a predictive-text usage rate of 30%.

1 Introduction

The telephone keypad, originally designed for the entry of digits, can also be used to enter text. Indeed, there are approximately one billion text messages sent per day from mobile phones (GSM Association 2003).

On a standard telephone keypad, several letters are assigned to each key. A letter can be entered by pressing the key to which the letter is assigned, and then choosing which of the several letters is meant by some method. The most common method is known as multi-tap, where the intended letter is obtained by pressing the key multiple times, depending on which letter is intended.

In the last few years, dictionary-based methods (DBMs) have been introduced to the market, and are now widespread. These methods work by matching an entered sequence of keystrokes to words in a dictionary. They purport to increase text entry speed by reducing the number of keystrokes required to enter each letter. In current cell phones, by far the most commonly installed dictionary-based method is marketed under the brand name T9® by America Online (AOL®). T9 has been adopted by a large number of manufacturers of cell phones. It is likely that nearly all of the exposure to a DBM reported in our study is in fact exposure to T9. T9 is marketed as a "predictive-text system". It guesses which word is intended by the user by matching a sequence of entered keystrokes against a dictionary of words. The match between sequences of keystrokes and dictionary words may be ambiguous. The DBM presents a selection of the words to the user in an order based on estimates of the relative probability of the words, and the user selects a word from the list. For instance, the keystroke sequence 22737 corresponds to 432 different letter sequences (with the standard 2/abc, 3/def etc. assignment of letters to keys). Of these, 13 are words in English: acres, bards, barer, bases, bares, baser, cards, carer, cares, caser, cases, caper, and capes. As we will see later, DBM methods are often called upon to enter words in a language different from the dictionary's target language. This can be a source of additional ambiguity. In this case, for instance, the Spanish word "abres" is ambiguous with the 13 English words. Proper names and acronyms are other sources of ambiguity. Typing mistakes can also lead a user have difficulties caused by ambiguity. For example, a user who intended to type "fares," but mistakenly pressed the 2 key rather than the 3 key to begin the word, would be presented with the list of words given above—none of which would be the intended word.

Despite these multiple problems, ambiguity may not be the major source of difficulties with dictionary-based methods. If the desired word is not in the dictionary, then it must be entered using multi-tap, typically after the list of presented possibilities is searched by the user. The time involved includes the time to realize that the dictionary has failed to find a matching word, to erase the entered word, to change to multi-tap mode, to re-enter the word using multi-tap and finally to leave multi-tap mode. A recent study by MacKenzie et al. (2001) showed that text-entry speed using DBMs is highly dependent on the proportion of desired words that are present in the dictionary. MacKenzie et al. showed that with 15% of words not in the dictionary, DBMs were slower for text entry than multi-tap. Studies by Grinter & Eldridge (2001) and Davis (1991) indicated that the proportion of non-dictionary words might be higher than 15%, but these studies were not directed specifically at measuring the proportion of non-dictionary words in short text messages. Thus, we sought to discover to what extent ambiguity, non-dictionary words, or other reasons, influence the acceptance of dictionary-based methods in the marketplace.

2 Methods

Interviews were conducted with randomly chosen people who were approached and interviewed on the spot in public places. Interviews were included in the data if the interviewee owned a mobile phone, and indicated clearly that they understood what a DBM was, and had used the system at least once. Each interview was videotaped for later analysis. No offer of payment was made. It is well known in the telecommunications industry that text messaging is most commonly used by teenagers and young adults, and that Finland is the nation with the highest adoption rate of new telecommunications technology. It can therefore be anticipated that young Finns would have among the greatest experience with, and highest adoption rate of, dictionary-based text entry methods. Thus, the bulk (150/230) of our interviews were conducted with generally young (15-30 year old) Finns. These took place in Helsinki and Turku, Finland. Additional interviews (80/230) were conducted with Catalan, Dutch, English, French, German, Italian, Norwegian, Spanish and Swedish speaking people in Barcelona, Ibiza, London, Munich, Paris, and Stockholm.

It is notoriously difficult to avoid experimenter bias in face-to-face interviews. To help avoid bias, it was explained to each interviewee before the interview began that the interviewer worked for a U.S.-based market research company charged with studying trends in SMS usage in Europe. It was felt that this would provide a neutral setting for questions that were in fact directed more particularly at the use of DBMs. Every effort was made to ask questions in a simple, matter-of-fact, non-leading manner, while at the same time providing a conversational setting conducive to spontaneous volunteering of information. It was found that an effective and non-leading means to draw interviewees into providing details was to simply repeat their statement as a question, and, where applicable, this technique was used throughout.

Interview Script. An interview was considered complete enough to include in the following analysis if it included a statement indicating some experience in using a DBM, and a statement as to whether the person used or did not use the method. Conditions permitting, additional information was sought, such as the make and model of phone used, the number of messages the person sent per day and the type of difficulties, if any, they had in using the system. A typical interview included several of the following questions:

Some phones have a built-in dictionary that predicts words as you write them, do you have that?

If yes: Do you use it?

If no: Have you tried it on somebody else's phone?

What kind of phone do you have?

Why (Why not) use the dictionary?

Do you put words into the user dictionary?

What fraction of the words which you use are not in the dictionary? Do your friends use it?

For users: How long did it take you to learn to use? Did someone show you how to use it?

For non-users: How many times did you try to use it? What happened when you tried to use it?

Sample Interview. The following is a transcript of a portion of an interview illustrating the interviewing technique. Interviewee comments appear in italics.

How many messages do you send a day?

Per day?

On average.

It varies, between 10 and 20.

Between 10 and 20.

Yeap.

OK, and what kind of phone do you have?

I have a Nokia 3210.

Nokia 3210, OK. The 3210 also has a dictionary system…

Yeap.

Called T9

Yeap.

Do you use that system?

No, I don't. I find it very frustrating.

Very frustrating. Why is it frustrating?

Doesn't work quick enough.

It's not quick enough?

Not quick enough, no.

Un huh. What do you mean, when you hit a letter it doesn't show up right away?

Well it does show up, but, you know, it suggests stupid words and stuff like that. I use usually slangs and shortened words and stuff like that, so I don't have really use for it.

3 RESULTS

Overall, DBMs are used by 47% (107/230) of interviewees. In the following, we attempt to tease out the various factors which lead individuals to either adopt or abandon the DBM.

Dependence of DBM Usage On Messaging Frequency. DBM usage as a function of reported number of messages sent per day is shown in Table 1.

Table 1: DBM usage as a function of reported number of messages sent per day. In each field, the proportion of usage is given, followed by the number of respondents in parentheses.

	Low	Med	High	Total
Msg/ day	0 - 2	3 - 5	> 5	
Female	0.35 (31)	0.36 (47)	0.54 (35)	0.42 (113)
Male	0.41 (41)	0.60 (42)	0.53 (34)	0.51 (117)
Total	0.39 (72)	0.47 (89)	0.54 (69)	0.47 (230)

The data for males and females combined show increasing use of a DBM as the number of messages per day increases. The difference between low and high is significant ($p<0.05$), according to the single-tailed *z*-test.

Proportion of Non-Dictionary Words. Of the 38 respondents to this question, 31 were DBM users, and 7 were non-users. When asked to guess what percent of words were not in the dictionary, some interviewees answered in terms of number of non-words per message. Table 2 shows the percentage of non-dictionary words calculated assuming different numbers of words per message. For each assumed number of words per message, the difference between the mean of the DBM user response and the non-user response is statistically significant ($p<0.05$, one-tailed t-test).

Table 2. Percentage of words not in dictionary estimated by users and non-users, for different assumed number of words per message.

	Words not in dictionary		
Words/msg	5	10	15
Users	14%	10%	9%
Non-users	21%	18%	16%

Users guessed in a range of percentage of non-dictionary words where a DBM should be faster than multi-tap, according to the findings of MacKenzie et al., while non-users guessed in a range where a DBM should be slower than multi-tap.

Summary of Comments by Users. There were 35 DBM users who elaborated on why they use the system. The top 10 reasons are given in Table 3.

Table 3: Reasons users cite for using a DBM.

Reason for using DBM	*Times mentioned*
Faster	26 (74%)
Easier	8 (23%)
Good idea	3 (9%)
Has different language databases	3 (9%)
Remembers a lot of words	2 (6%)
Surprisingly often, gets right word	2 (6%)
Works well	2 (6%)
Useful/helpful	2 (6%)
Checks spelling	2 (6%)
Stores abbreviations	1 (3%)

Thus, the major reason for using the DBM is that it is faster than multi-tap, and we might expect that, symmetrically, non-users would cite "it's slower" as a major reason for not using the DBM. We will see below that this is not the case. Before turning to the responses of non-users, let us consider the problems encountered by users, and their means of solving them.

Problems with the dictionary. Users were typically asked, "Do you have any problems with the dictionary?" 53 users responded with details. The top 10 responses are listed in Table 4.

Table 4: Responses from DBM users to the question "Do you have any problems with the dictionary?"

Remark	Times mentioned
Words not in dictionary	31 (58%)
No slang words	11 (21%)
Gives wrong words	9 (17%)
Can't mix languages	4 (8%)
Names not in dictionary	3 (6%)
Doesn't know word endings	3 (6%)
Annoying	3 (6%)
Need to proof read the text	3 (6%)
Have to switch language databases	2 (4%)
No swear words in dictionary	2 (4%)

By far the dominant response refers to the general problem of non-dictionary words. Additional responses cover classes of non-dictionary words, such as slang or swear words, or names, or words from other languages. As we will discuss below, "wrong words" could be due to words not in the dictionary, or ambiguity, or typing and spelling errors, or still other sources. Among users, "annoying" is most strongly related to non-dictionary words. One user, who estimated that 30% of words were not in the dictionary, stated,

"Who's a happy customer anyway, if 30% of words are not in the dictionary? It gets most of the words right, some of them wrong. Still, it pisses you off when it gets it wrong, but still, it's a good system."

Another user, when asked to estimate the probability that a word is not in the dictionary, replied,

"Not so often. But when you're used to something that works and suddenly it doesn't work, you get annoyed."

Adaptation to DBM problems. Among users, the problems reported with DBMs appear to have induced adaptations to better deal with the system. The easiest adaptation is to ignore the problem,

"When I am tired, I write the message in T9 without correcting, and the words are not correct, but I think my friend will understand."

In other cases, users find work-arounds:
"I'm Swedish, but I use Finnish words sometimes, and then it can't find the words. So I have to write very good Swedish. It can't find the Finnish words, or English words."
Here the user has compensated for the lack of Finnish words in the DBM by modifying her normal writing so that it includes only Swedish words. A Finnish user,
"Compound words are the problem... then I have to decide whether to discard the word or not, and it might even result that I change my mind and say something else to avoid it."
This user is also devoting a high level of cognition to selecting words not for their ability to express his thoughts, but according to the likelihood that the dictionary will accept them.
"Sometimes I have to switch it off because it doesn't recognize the words, since it's not strictly Finnish. Slang words, shortened words. It doesn't work usually on the kind of words people normally use to save space. But of course you can teach the machine, but it simply takes too much time. Now and then I switch it off, but I do tend to use it because it is more rapid."
This user is experienced enough with the system to be able to largely anticipate, before a word is typed, whether the word is likely to be in the dictionary or not. As there are tens of thousands of words in the dictionary, this adaptation is unlikely to be memorization of words in the dictionary. Rather, it involves some higher-level classification of words, into types likely to be accepted by the DBM, and types likely to not be accepted.

Being able to anticipate non-dictionary words means being able to avoid the laborious process of writing the word first, then finding that the dictionary doesn't know it, and then having to write it a second time using multi-tap. The advanced knowledge makes the non-dictionary word less of an obstacle. More generally, knowledge about "*how the machine thinks*," as one user puts it, seems to be a hallmark of the user, as opposed to non-user, profile. There are exceptions, however, and non-use cannot simply be attributed to lack of training. One non-user explained,
"*I saw how it works, and I thought, 'I'm never gonna use it'.*"

Adaptation to a text entry method has also been observed in the case of multi-tap (Grinter et al. 2001).

User dictionary. Some DBM implementations include a user dictionary that allows users to enter new words using multi-tap, and then makes them available the next time they are needed. For some users, the user dictionary palliates the effect of non-dictionary words, though not all users of the DBM use the user dictionary. Those who do often point to it as a key feature of the DBM. Indeed, a good number of the DBM users stated immediately after mentioning that some words are not in the dictionary, "*you just add them* [to the dictionary]." As an indication of the level of dependence user might have on the user dictionary, some report filling up the user dictionary. An example,
"*The memory gets full very fast. It's very small that memory.*
Interviewer: "*Have you filled it up*?"
Yes, with my own words."
Another asks the interviewer, "*Do you know how many words it can keep? I think it can't keep so many words*."

Some users attribute their speed in using the DBM to complete training of the user dictionary, as illustrated in this exchange between a user and non-user.
Non-user: "*It's pretty good, but it's so slow...*".
User (interjects): "*You have to teach your mobile phone, and learn your language. [...] When your mobile phone knows your language, what <u>you</u> have, then you can write fast.*"
Interviewer: "*How long have you had your mobile phone?*"
User: "*A couple of months, and now I can write fast. I taught my phone, and now I can write fast.*"

Summary of Comments by Non-Users. 102 non-users gave answers when questioned as to why they did not use a DBM system. The 10 most frequent responses data are given in Table 5.

Table 5: Top 10 responses by DBM non-users to the question: “Why don’t you use the dictionary?”

Response	*Number of responses*
It is /difficult/complicated/	45(45%)
It gives /wrong/unwanted/strange/stupid/ words	36(36%)
I don't /know how to/can't / use it	29(29%)
It’s too slow	18(18%)
I’m /familiar/comfortable/with the old system	13(13%)
/Dictionary too small/unknown words/	13(13%)
It doesn’t work	13(13%)
It doesn’t have slang words	9(9%)
I’ve no /use/need/ for it	8(8%)
Can’t use /short words/abbreviations/	7(7%)

It is too difficult. Consider, for instance, the complaint that the DBM is difficult or complicated. Sometimes this complaint is expressed in a fairly pure form,

“*It’s getting too complicated for me. I prefer the old way.*”

In other cases, the system is “complicated” because it “gives wrong words,”

“*Because it’s so complicated, it gave me wrong words. Usually, because I use those words that kind of word that they don’t find that, not ordinary words,*”

or because it’s hard to use,

“*I don’t know how to use it. Have tried to use it, but never got into it, tried about 5 times. I get frustrated because I don’t know how to use it. It’s very difficult to get into it. It’s much quicker to write with my own words.*”

It doesn't work. Some non-users are of the opinion that the system simply doesn’t work. Some felt that the cause of failure might lie with themselves, “*It doesn’t work. Or I can’t use it. Either/or,*” while others placed the blame resolutely on the phone itself,

“*My mom has it and she thinks it’s not that good, that it doesn’t work properly. The machine guesses the wrong words. I suppose there is something wrong with the phone itself. She should get it fixed.*”

Some non-users explained what they meant by “not working,” giving wrong words in the example above, being slow in the example below,

“It doesn’t work, or it doesn’t speed my writing. It’s much slower. I haven’t bothered using that. At the beginning it was so slow. You’re better off without the dictionary.”

It gives wrong words. Some users simply say that words are “wrong,” and are at a loss to explain it:

“*Sometimes I have to put words again and again because they go wrong.*”

“There are wrong words coming and going away and I don’t know why.”

However, the complaint of “wrong words” is often linked to other aspects, such as being complicated or slow:

I don’t like it. When you try to write a word that’s not in the memory system then it predicts another word, it makes it complicated.”

Or,

“*It has so many wrong words, and it’s so slow.*”

A “wrong word” can arise from a variety of sources (acting singly or in combination). These include words which are in the dictionary, but which are ambiguous with other words, the others being of higher probability, and so presented by the DBM first in the list. For example, if the word “tie” is desired, the word “the” will be shown, since “the” and “tie” are typed using the same keystroke sequence, but “the” is more probable than “tie”. One non-user noted,

“*Sometimes I start to use it and then it doesn’t work and then I change. It gives you 3 or 4 options and you never find your word.*”

A word could also be “wrong” if it is not in the dictionary, and the system presents a guess that appears random to the user. Additionally, any typing error or spelling error made by a user is likely to produce a “wrong word” as the DBM does its best to match the incorrect input sequence to a correctly spelled but (by definition) unwanted word. Either of the latter two reasons, non-dictionary words or spelling/typing errors would be most consistent with the user studies of MacKenzie et al..

Dictionary too small. Some non-users, like many users, are able to specifically point out classes of words which are not likely to be in the dictionary, such as surnames, *"It recognizes all the names of the Nokia bosses, but not all the names, even common names,"* or swear words,

"And other times we need some terrible words, and they're not in the dictionary," but, generally, non-user responses more generically point to wrong or missing words, rather than to specific classes of non-dictionary words.

Summary. Non-dictionary words make the system appear not to work, and, since they need to be handled in a special way using a secondary system (multi-tap), they make the system awkward to use and hard to learn. Non-dictionary words are complicated and slow to deal with, and can cause annoyance and frustration. Correcting the "wrong" response of the system when a non-dictionary word is typed can take additional time, and require the user to master the keystroke sequences required to effect the correction. Many of the varied comments of non-users are manifestations of problems directly related to desired words not being present in the dictionary.

4 Modeling The Discovery Period

Fifty non-users answered the question: "How many times did you try to use the dictionary-based system before you gave up on it?" The distribution of responses is shown in Table 6.

Table 6: Number of times non-users attempted to use the DBM before abandoning it. The mean number of tries was 3.6.

Times Tried	*Number of responses*
1	11 (22%)
1-2	15 (30%)
3-5	15 (30%)
5-10	8 (16%)
> 10	1 (2%)

Assuming a trial consists of entering just a few words, perhaps 5 to 10, then the average non-user tries to enter 18 to 36 words before abandoning the system. This is the trial period during which the DBM has an opportunity to convince potential users that it has real utility for them. We will call this trial period the "discovery period." Understanding the user experience during this discovery period is vital for understanding product adoption.

To study the discovery period more quantitatively, we obtained data from a user study by MacKenzie et al. (2001) in which 10 subjects typed for a total of 10 hours each, using multi-tap. We used the first 25 words typed by each subject to analyze their discovery period. Thus, our input data consists of actual keystrokes made during the discovery period *of multi-tap.* We then converted these keystrokes to those that would have been made by the same users *had they been using a DBM to enter the same words.*

Non-dictionary words. We configured the DBM simulation so that the least common 15% of words from the DBM dictionary. Entry of non-dictionary words was simulated by allowing the user to enter the word using the DBM, and then immediately reentering the word in multi-tap mode. It is to be emphasized that we are modeling naive users, who do not yet possess sufficient knowledge of the dictionary system to be able to anticipate that a word will not be in the dictionary. And yet, we do not include any time for the user to realize that a non-dictionary word has been encountered and to decide how to handle that event. That is, we assume that the beginning user has read the users manual and knows how to properly manipulate the DBM. This is very favorable to the DBM, and our results should be seen as a lower bound for the time required to enter words using a DBM.

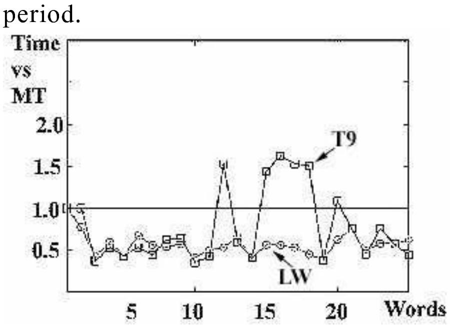
Error correction. Handling the effect of typing errors is an important and subtle aspect of the simulation, and our method follows that of MacKenzie et al. Typing errors made when using DBMs are typically not detected immediately. This is because the screen often displays a confusing and unreadable sequence of letters during word input. Users of the T9 DBM system who have read the users manual may encounter additional delays in recognizing that an error has been made, since these manuals instruct the user to not to look at the screen while the word is being entered, and to only look at the end of the word. Some T9 implementations emit a beep as soon as the user has typed a sequence of keys which no longer matches a word or partial word in the dictionary. This beep may alert the user that something has gone wrong, and in these simulations, we made the favorable assumption that a beep is emitted under these circumstances in our simulations. On average in the McKenzie data, the beep occurred on the penultimate character of the word, which in practice saved the user two keystrokes per typing error (MacKenzie et al. 2001).

LetterWise. To allow more complete comparison with the results of McKenzie et al., we also included a simulation of the discovery period of the third text-entry method they studied, called LetterWise. LetterWise is an alternate method of text entry that does not use a stored dictionary of words. Instead, a small database of prefix information is used to disambiguate user keystrokes. In all respects, including handling of typing errors, our simulation of LetterWise follows the approach of McKenzie et al.

The per-word KSPC. The per-word keystrokes per character (KSPC) is the number of keystrokes required to enter a word, divided by the number of letters in the word. When typing errors were made, all keystrokes, including backspaces, if any, were included in the total count on keystrokes for that word. When simulating the DBM we ignored any key presses needed to scroll through sets of ambiguous completed words, a procedure which is favorable to the DBM.

Figure 1 shows the initial experience for a typical subject. Although the KSPC for most words typed using a DBM was less than the KSPC for the same word typed using multi-tap, some words took more keystrokes in the DBM than in multi-tap.

Figure 1. KSPC (relative to the KSPC for multi-tap) for each of the first 25 words for a representative subject. Spikes above the multi-tap baseline (1.0) indicate words that took more keystrokes to enter than were required using multi-tap. For all ten subjects the average number of spikes above the multi-tap baseline was 4.9 (standard deviation 2.9) during the discovery period.

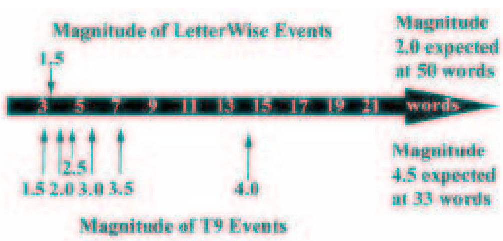


As seen in Figure 1, a DBM with non-dictionary words exhibits bi-modal behavior. Words which are not in the dictionary, or words in which a typing mistake is made, may take longer to enter in the DBM than in multi-tap, sometimes much longer. A user who encounters many non-dictionary words or who makes typing mistakes during the discovery period may conclude that the DBM is much slower than multi-tap, and may therefore may abandon it. A user who encounters fewer non-dictionary words, or makes fewer typing mistakes may arrive at the opposite conclusion.

Note that LetterWise, while slightly slower than the DBM on some words, did not produces spikes beyond the multi-tap baseline. In general, it is much more consistent in its behavior than a DBM.

Schematically, we can summarize by considering the expected time to "earthquakes" of various magnitudes. That is, how many words will a beginning user enter until they encounter a word which takes 1,2,3… times as many keystrokes to enter using T9 or LetterWise as it would to enter using multi-tap? It can be anticipated that if a beginning user encounters a word which takes much longer to enter using the new system than it would have using their familiar multi-tap system, then they will be put off, and unlikely to explore further to gain mastery of the new system. Figure 2 shows a time scale, measured in number of words entered, and points at which events of various magnitudes can be expected.

Figure 2. Expected time to "earthquakes" of various magnitudes, using T9 and LetterWise. The magnitude of the event is the ratio of the number of keystrokes required using the predictive system compared to multi-tap. Thus, a magnitude 2 event is a word which takes twice as many keystrokes to enter with the predictive system as it does with multi-tap.



5 DISCUSSION

The study by MacKenzie et al. (2001) showed that text-entry speed with dictionary-based methods is highly dependent on the fraction of words in the dictionary. In this present study, significant numbers of users and non-users reported that their main difficulty with dictionary-based methods is related to words not in the dictionary.

While users clearly pointed to non-dictionary words as being the main problem they had with the DBM, non-users often reported that they found the DBM difficult to use, without direct mention of non-dictionary words. Upon closer examination of their comments, their usability problems can often be traced to difficulties in handling non-dictionary words. In particular, the main complaint, that the dictionary is hard to use, is related to the unpredictable nature of the response of the DBM to the entry of non-dictionary words. The situation is similar to that studied by Wiener (1989), in his case, user response to automatic flight control systems. When confronted with the response of the system to unusual events, the most frequent questions asked by pilots were, “What is it doing?,” “Why is it doing that?”. To be comfortable with the system, users need to build a mental model of the behavior of the system. In the case of the DBM, the mental model must include an understanding of the types of words likely to be included, or not included, in the dictionary. This is fairly sophisticated linguistic knowledge, which takes time and motivation to acquire. However, prospective users usually gave the DBM but a few chances to demonstrate its utility to them, entering a handful of trial phrases. This amount of time is often insufficient for users to grasp the subtleties of the system, and to discover adaptations to its short comings.

In an attempt to quantify the effects of non-dictionary words on initial user experience, we conducted a simulation study based on previous work of MacKenzie et al. We took data from their user study in which subjects typed using multi-tap. We then built a simulation of a dictionary-based method from these data, and also a simulation of a prefix-based method which does not use a dictionary. These simulations revealed that dictionary-based methods have inconsistent behavior. When a word is in the dictionary they perform significantly better than multi-tap, but when the word is not in the dictionary, or when a typing error is made, they perform significantly worse than multi-tap

The bi-modal behavior of DBMs is likely to play a determining role in the initial and continuing user experience, as consistency is an important property of comfortable and discoverable user interfaces. It is ventured that the mass acceptance of text entry systems, and new technology in general, is not determined by average-case behavior of the system in the hands of an expert user, but by the worst-case behavior in the hands of a novice user.

References

Davis, J. R. Let your fingers do the spelling: Disambiguating words spelled with the telephone keypad, *Avios Journal 9* (1991), 57-66.

Grinter, R. E., and Eldridge, M. A. Y do tngrs luv 2 txt msg? *Proceedings of the European Conference on Computer Supported Cooperative Work – ECSCW 2001.* Amsterdam, Kluwer Academic Press, 2001.

MacKenzie, I.S., Kober, H., Smith, D., Jones, T., and Skepner, E., LetterWise: Prefix-based Disambiguation for Mobile Text Input *Proc UIST* 2001, 111-120.

Wiener, E.L, Human factors of advanced technology (“glass cockpit”) transport aircraft *(NASA Report 177528).* Moffett Field, CA: Ames Research Center (1989).

http://www.gsmworld.com.

Döring, N. (2002). "Kurzm. wird gesendet" - Abkürzungen und Akronyme in der SMS-Kommunikation. *Muttersprache. Vierteljahr-esschrift für Deutsche Sprache* 112(2), 97-114. ISSN0027-514X..www.nicola-doering.de/publications/sms-kurzformen-doering-2002.pdf