Publications: One Finger, 10 Keys and 6,000 Characters

Carl Kay

Globalization Insider - March 19, 2004

http://www.lisa.org/archive_domain/newsletters/2004/1.3/kay.html?access=manioca

The Dilemma for Japanese Cell Phone Users

Gangster movies notwithstanding, most Japanese still have ten fingers. But when using a cell phone moving around in Tokyo, nine of those fingers are otherwise occupied holding the device itself (four of 'em) and the strap one clings to for life (the other five) in the sardine-can packed subway. That leaves people with one finger and ten keys to deal with 6,000 characters! Carl Kay explains how cell phone manufacturers (and users) are rising to the challenge.

The Japanese have gone from typing 6,000 characters with ten fingers on 100 keys to doing it with just one finger and only ten keys. According to Naruhito Tanaka of Omron Software, "In Japan, about 10 million people touch-type on PC keyboards, while over 30 million send mail from their cell phones. So the cell phone keyboard is the defacto standard now." (2002, Japanese). When I wrote "Japanese Software Wars" for Language International (9.3:10–11, 47) back in 1997 about the linguistic and technical issues of inputting Japanese text using a PC keyboard, and the related battle between Microsoft and Japan's Justsystem to dominate the Japanese PC desktop, I never imagined the reality of 2002 described by Tanaka. Having once tried to explain the challenges of typing 6,000 characters with only ten fingers and the 100 keys of the PC keyboard, here I'd like to describe the emerging technical, business and cultural trends related to the challenge of typing the same 6,000 characters with just one finger and only ten keys.

Lack of personal space in Japan makes text an important means of communicating privately.

Harmony Through Communication

Gangster movies notwithstanding, most Japanese still have ten fingers. But when using a cell phone moving around in Tokyo, nine of those fingers are otherwise occupied holding the device itself (four of 'em) and the strap one clings to for life (the other five) in the sardine-can packed subway. But the one free finger has an important role to play. (Recently, two-thumb operation has begun as the number of function keys has increased on Japanese phones. Still, this assumes that you are sitting on the train and thus have both hands free.) Good relationships in Japan require harmony maintained by frequent communication to confirm the other person's response. Short text messages serve this function well.

Phone conversations can perform this social function too, but the lack of personal space in Japan makes text an important means of communicating privately. (In Japan, 3M [Japanese] is now selling a cell phone screen privacy filter that limits the view of the screen to the owner's straight-on perspective. So you may be squeezed on all sides by a mass of commuting flesh, but at least none of those ever-so-close eyeballs will be able to read your messages.) Also in the subway, which many people ride several times during the work day as well as for commuting, voice calls get cut off frequently. The rhythm of the train journey fits text messaging much better. You can compose and read messages when out of range in the tunnels and send and receive them when the train stops at stations, which all are covered by Japan's wireless carriers.

So just how does one type 6,000 characters with one finger?

The Cell Phone Challenge

So just how does one type 6,000 characters with one finger? First, as with PC input, a Japanese writer must input characters from a syllabary of about 50 phonetic characters called kana that represent the sounds that make up all Japanese words. On the cell phone keyboard, typically the number key 1 represents the A row of A, I, U, E, O; the 2 key represents the K row of KA, KI, KU, KE, KO; the

3 key the S row of SA, SI, SU, SE, SO and so on, with the other numbers representing the remaining consonants in rows which all follow the same a, i, u, e, o pattern.

The first real challenge is to decide how to assign each single number key to one of the five different sounds included in its "row." Among the early solutions to this problem was the multi-tap method, where the same key is repeatedly pressed to call up the a, i, u, e, o sequence in succession until the correct character is reached. Thus, a word like ko-do-mo 'child' requires fifteen key taps. Another primitive system, which surprisingly still retains some loyal users in Japan, is the poke-beru or pocket bell pager text method input, whereby numbers are input to indicate a character's position in the consonant row. For example, the phonetic character TO occurs fifth in the T-row of the syllabary that consists of the five characters TA, TI, TU, TE, and TO. To input TO, you tap the number key which has TA on it (usually the number key 4) and then tap the number key 5, indicating that the character desired is the fifth one in the T-row.

Many phonetic sounds can specify up to ten or even twenty possible characters.

With Japanese, inputting the phonetic sounds is just the first input step, and in some ways it is the easier one. Next, the writer must convert some of these syllables into characters chosen from about 6,000 Chinese-derived kanji characters. This task is complicated by the absence of word breaks in Japanese text. The conversion software, not infrequently, will guess wrong about where you intend word breaks and will offer a string of Japanese words–all perfectly correct word by word, but which are utter nonsense when strung together. A simple example, if English had no word breaks, would be if you typed hisid, intending to write hi Sid, and the software displayed from its dictionary of valid words the expression his id.

Due to the lack of word breaks, you often have to spend time (and keystrokes) parsing the sentence manually to clarify which sounds are to be considered together as words, so that the software can perform the kanji conversion operations on the words you intended. This step can be annoying enough on a PC keyboard, but is even more so on a cell phone where you have fewer keys and fingers to specify the rearrangement.

The next big complication is that many phonetic sounds can specify more than one of the 6,000 kanji characters (Japanese is full of homonyms, so some phonetic sounds represent up to ten or even twenty possible characters). When the string of phonetic characters has finally been correctly parsed into word groupings, Japanese typing many times requires further manual steps, often through some kind of pull-down menu, to override the software's default "guess" at the kanji you intend for a given phonetic input. For example, the characters that spell the name of my neighborhood in Tokyo can also mean 'normal,' 'spotlessly clean,' or 'political circumstances,' each of these words being a homonym rendered by totally different characters. My software has a "learning" function and now "knows" that I usually am typing my address with those particular phonetic characters and now offers my neighborhood name as the first choice. I can override the software if I happen to mean one of the other characters. The better conversion engines also have some user-independent, built-in contextual awareness so that, for example, the homonym with the medical-related meaning is the one automatically offered as the first choice in a text where other medical terms appear frequently. In any case, obtaining the correct character as frequently as possible as the first choice offered is critical for ease of use and productivity.

Justsystem vs. Microsoft on the Desktop

Toshiba pioneered kana→kanji conversion engines for mainframe and mini-computers in the late 1970s, but it was Justsystem, a small company from Japan's remote southwestern coast, that surged to rapid growth in the 1980s by creating the first widely accepted IME (input method editor) for PCs. Just's ATOK program was the standard IME bundled in NEC's proprietary 9800 Series of PCs , bundled with Just's highly profitable word processing program, Ichitaro. This configuration ruled the desktop in Japan during that decade.

Then in the early 1990s, Wintel crashed the PC party in Japan, and now Dell has supplanted NEC, Word has eclipsed Ichitaro, and Microsoft's own IME comes standard with most PCs. Some loyal ATOK users still go to the extra trouble and expense to substitute ATOK for Microsoft's IME. Justsystem has fought hard to keep ATOK at the cutting edge (it is now in version 17), boasting of such functions as the ability to convert accurately the non-standard phonetic spellings of Japan's many regional dialects. But on the desktop, the game is basically over. Microsoft has won, and Just has turned to areas not (yet) dominated by Microsoft, such as enterprise knowledge management, in an effort to revive growth.

The situation is quite different in the mobile space.

The Mobile Space Is Wide Open

However, the situation is quite different in the mobile space. Microsoft is absent from the market, both in the West and in Japan. Multi-tap was clearly not the final answer, but on the cell phone keyboard it was up to a different Seattle-area company to find a better way. Tegic, a company founded in the mid-90s with the goal of making PCs more accessible to physically challenged computer users, later applied its technology to pioneer "predictive input" on cell phones with such success that the company was acquired by AOL in late 1999. Explains Ray Tsuchiyama of AOL Japan, who pioneered Tegic's move into Asian markets, "With only the ten keys of the cell phone, we are all disabled to some degree. Anything that reduces the number of operations is a big help." [February 2002, phone conversation with the author.]

Predictive input software tries to guess ahead what the user intends to input, based on (1) what has been input so far, (2) an extensive analysis of what that user and thousands of others have done in the past, and (3) the range of possible words in a language. You can imagine how predictive text works in English by thinking of how the list of valid completions narrows down as one inputs the letters A (thousands of possible words), then B (down to a hundred or so), then O (a few dozen) and U (only a few).) Good predictive software would probably offer the word about fairly early as a high probability choice, though sometimes the user might mean abominable.

A significant amount of time can often be saved by using predictive text instead of multi-tap. Successfully executed, this greatly speeds up the cumbersome process of text input on cell phones, since each word automatically input by the system potentially saves so many strokes. As such, it applies to cell phone text input in any language, not just Japanese. In fact, Tegic first found acceptance in European markets beginning in 1999.

"The algorithms for European languages were easier than for Japanese, so Tegic's product T9 was completed for European languages before it was finished for Japanese," recounts Tsuchiyama. "Cell phone text messaging had just started in Europe, and people were using the multi-tap method, pressing the 2 key three times to get through A and B to input C and the like. T9 offered a clearly better input method, and since nobody had been using multi-tap for very long, there was not much resistance to switching." Tegic found acceptance among European cell phone manufacturers and among Japanese phone manufacturers producing models intended for export to Europe.

Oyayubi ("Thumb") Culture

But inside Japan, a strange thing happened. Early cell phone adopters, mainly young people and especially young women, proved so adept at multi-tap typing that an entire sub-culture called "oyayubi (thumb) culture" flourished, with accompanying fashions and hairstyles that influenced cutting edge trends worldwide for several years. All the dexterity gained with training from childhood on chopsticks and the abacus, along with years of having no other choice, made people more comfortable with multi-tap than AOL/Tegic and others had guessed.

Undoubtedly, the multi-tap method gained a foothold partly because of the short, simple and shallow nature of much of the text being input. According to Tim Clark (personal e-mail communication , February, 2004) publisher of the Japan Entrepreneur Report and noted expert on Japan's mobile culture, "It is not uncommon for 13-year-old girls in Japan to send and receive 100 mail messages per day via mobile handsets, mostly in the form of protracted, line-by-line conversations about school, meals, boys and extracurricular activities. Older teenagers, college students, and even working adults are nearly as enthusiastic about mobile mail, and while the content of their communications may be less juvenile, it is still primarily personal rather than professional. For these users, messages consist primarily of appointment requests or confirmations, or short questions or reminders sent to friends, colleagues, spouses or children."

The range of vocabulary, and thus the range of kana→kanji conversions needed, remained limited, which minimized the motivation to go beyond multi-tap to a more sophisticated method that would facilitate more complex text input. Cutting a 90- second input to 40 seconds may not matter so much if the typing is done riding between train stations, when there is nothing else to do anyway. Also, since Japanese kana represent syllables rather than letters, it only takes multi-tapping two kana characters for, say, the word ha-na 'flower.' However, the same operation requires specifying six letters for the equivalent word in English and five in French (fleur). Therefore, multi-tap may be less of an innate handicap in Japanese.

Enter Predictive Text Input

Yet, as cell phone mail usage spreads beyond the young crowd, users need a reasonably speedy input method that relies less on physical dexterity. Using the phone for business and other less casual purposes can require more complex messages. The physical limitations of the cell phone device as a keyboard call out for a "next level" solution such as predictive text input. The complexities of the Japanese language slowed down development of algorithms for predictive text input of the language, but AOL/Tegic and others are now bringing predictive text to the Japanese market. Omron's Tanaka says, "Raw kana→kanji conversion accuracy is getting good everywhere, so usability [in the form of predictive text] will become the key differentiator going forward." Which conversion choices are offered by the predictive input system, which less likely choices are suppressed, how the screen is arranged, the timing of the offered choices, the ease of making a selection—these will be the points that users judge when selecting an input method and will appreciate as they type their mail on crowded trains.

If a Japanese predictive text system has correctly predicted the word(s) intended, it by default automatically offers the correct kanji as well. Thus, in Japanese, predictive input not only reduces the number of taps required to obtain the desired kana characters; in many instances, it eliminates the very need for the subsequent kana→kanji conversion step. Thus, there is a kind of double speed gain that makes this trend extremely significant as cell phones become securely established as the primary text messaging medium for potentially a hundred million people in Japan.

Current Offerings in Predictive Text Input

Just's predictive text product, APOT, leverages its ATOK heritage and wins praise for its accurate handling of regional dialects, conversational-style text and references to currently popular people, places, trends, etc. which are even more common in cell phone messages than in desktop-based mail.

Another predictive text input that currently has high market penetration is Omron's Advanced Wnn (pronounced "Un-ne"). Omron claims that its prediction algorithm offers the correct choice between nine candidates 90% of time (in other words, in one screen view and selectable by pressing one key). Like Justsystem, Omron claims its system is tuned to recognize regional dialects, and even male and female style expressions, which can be quite different in Japanese. Amazingly, the company even claims the system is tuned to recognize certain expressions used by popular animated characters from Japanese TV, which are used frequently in email conversations.

A current catalog from the Japanese mobile carrier Au (KDDI[Japanese]) lists the supported input method(s) of each of its fifteen current handset models as one of the specifications buyers can use (along with battery life, camera pixels and degree of GPS support, etc.) to select a phone. Predictive text systems by Just, Omron, Toshiba and Sony are available, depending on the Au phone model. Other carriers offer these systems, as well as AOL/Tegic's T9, in addition to an OEM system from Fujitsu and even one from the homebuilder Misawa. A good predictive input method is becoming a selling point in the handset market, and phone manufacturers and input method developers are working feverishly to create popular configurations.

AOL/Tegic's T9, now included on all NEC i-mode 3G handsets and on a few models by other manufacturers, relies on the traditional Japanese a, i, u, e, o order of the kana characters, but adds intelligence that greatly speeds up input. To type Matsui, you type 7, which on the Japanese phone has MA printed on it. The T9 system, like earlier systems, represents the M-row in the a, i, u, e, o system, namely the syllables MA, MI, MU, ME and MO; then 4 for TA which covers the T-row, TA, TI, TSU, TE, TO; and followed by 1 for A, which covers the A-row, A, I, U, E, O. From the three keystrokes, which theoretically could be meant to specify any of 53, or 125 different possible combinations of syllables, the software's algorithm quickly narrows the selection to a few (or in other cases only one) that spell a valid word, and the user simply selects one. For example, the 7-4-2 sequence above might produce Ma-tsu-i, but could also mean another name, Mi-tsu-o, as well as other words. The longer the selection process, the fewer the choices that remain valid, as each letter is specified, eliminating formerly possible completions of the user's input. In contrast, the same word Matsui requires six taps using the old multi-tap method: one tap on the 7 key for MA, three taps on 4 to cycle through TA and TI to get to TSU, and two taps on 2 to get through A to I.

Omron's Advanced Wnn and Just's APOT systems use multiple methods to anticipate frequently used words to offer the desired completion from just one or two characters. Similarly, a word or even a part of word can bring forth a prediction of a whole phrase. For example, inputting O-HA quickly produces O-HA-YO-U, followed by GO-ZA-I-MA-SU. (Ohayou gozaimasu is 'good morning' in Japanese.)

The memory space in a cell phone is very limited, and the dictionary for text input competes for space in the design process with other functions such as photo editing. Advanced Wnn in its basic configuration takes up 828kb ROM (61kb for the engine and 767kb for the basic dictionary), plus 73kb of RAM (including 30kb for the engine and 32kb for the learning function). Thus, it is impossible to build a huge dictionary to cover all possible content areas, as you would on a desktop system.

One current solution to this limitation is for users to download (1) specialized small dictionaries in their fields of work or for their hobbies, (2) place name dictionaries focused on their regions, (3) along with other narrowly focused subjects, to enable them to customize the limited dictionary space available. Also, since much efficiency is gained by the "learning" that the software does to anticipate each user's likely input choices, provisions are made to upload the file that holds that accumulated learning to a server so that it can be reinstalled whenever the user switches handsets. (Anything that creates a disincentive to upgrade handsets frequently is, of course, anathema to handset manufacturers.)

It took the severe constraints of the cell phone keyboard, and the lure of a huge market, to push innovation in Japanese text input on an order of magnitude higher than was achieved in the battle for the desktop. Still, the benefit of some of these new input systems born from mail desperation in Tokyo's subways may be brought to PC users in Japan. Perhaps Omron, Justsystem or another manufacturer will apply these advances to the PC to rise up and re-challenge Microsoft for control of the Japanese desktop. Or maybe for their next breakthrough, Sharp, Casio, Panasonic and Sony will make cell phone keys even smaller and fit 100 of them on a phone, and the oyayubi tribe of Shibuya will figure out how to type rapidly using those as well. Whatever direction technology takes, we can be sure that the innovations for inputting Japanese will continue to surprise.