Near Eyes-Free Chauffeur Computer Interaction with Chording and Visual Text Mnemonics

Frode Eika Sandnes
(Faculty of Engineering, Oslo University College, Oslo, Norway
frodes@iu.hio.no)

Yo-Ping Huang
(Dept. Electrical Engineering, National Taipei University of Technology, Taipei, Taiwan
yphuang@ntut.edu.tw)

Yueh-Min Huang
(Dept. Engineering Science, National Cheng Kung University, Tainan, Taiwan
huang@mail.ncku.edu.tw)

Abstract: Modern cars are equipped with advanced technology requiring cognitively complex operation that is reliant on the user’s visual attention. It is therefore hazardous for drivers to operate such devices while driving. In this paper a user interface interaction style for in-car user interfaces are proposed. Users interact with the in-car computer using three chording keys and chording pattern sequences are derived based on visual mnemonics. Cases are illustrated for an in-car multimedia system, a mobile phone and a GPS-navigation system. Experimental results demonstrate that the technique is easy to learn, efficient to use and require low visual attention.

Keywords: Human computer interaction, mobile text entry, chording, ubiquity, limited visual feedback, spatial mnemonics, in-car user interface.

Categories: H.5.2

1 Introduction

Cars are emerging as increasingly complex technological environments for drivers. Cars are no longer simply controlled using pedals, steering wheels, and the gear stick. In addition to the various switches on the dashboard that control essential functions such as lights and window wipers new functions have emerged. Drivers often learn to use these simple controls with limited visual attention allowing drivers to focus their eyes on the road while driving. The human motor memory is a powerful mechanism that allows humans to accurately memorize where in space an object is located. The user occasionally needs to look at the indicators on the dashboard such as the speedometer, fuel level and warning indicators.

However, many new ubiquitous and pervasive technologies have emerged in in-car environments including mobile phones. It has been found that approximately 8% of drivers occasionally operate mobile handsets while driving [Walker, Williams, and Jamrozik, 2006]. Activities range from answering calls [Treffner and Barrett, 2004] to making outgoing calls, and even indulging in hazardous activities such as reading and responding to text messages while driving. Several studies have looked at how a
secondary task such as mobile phone use affect the primary task in terms of driving performance [Reed and Green, 1999; Salvucci, 2001; Tsimboni, Smith, and Green, 2004].

In-car multimedia systems has also become commonplace. A multimedia system may include a traditional FM radio, a CD-player, an mp3-player, and DVD-players. Although the driver cannot watch DVD-movies while driving, children may stay less bored during long journeys when watching movies. However, drivers often focus their attention tuning to find the right radio station or navigating a CD or mp3 music collection.

Finally, personal GPS navigation systems have become affordable and commonplace. Most of these provide speech output allowing the driver to follow a set route by listening to synthesized spoken instructions. However, the input systems on such devices often depend on visual feedback.

One important challenge in in-car user interfaces is to reduce the visual demand [Burnett and Porter, 2000]. Most work on in-car user interfaces has focused on speech based interaction. Other studies have looked at vibrotactile feedback [Jan, Van, Hendrik, and Van, 2004]. In this study an in-car input technique based on chording is proposed. The system is easy to learn and is not reliant on visual attention.

1.1 Chording

Chording has traditionally been used for text input and has been around for a long time [Noyes, 1983]. Users input chords in a similar manner in which musicians hit chords on a piano. A chording keyboard therefore only has a few keys compared to traditional QWERTY-keyboards. Each chord has a unique meaning. Typically there is a chord associated with each alphabetic symbol [Lyons, Starner et al., 2004]. Chording keyboards require little visual attention as there is no need visually search for keys once the hands are correctly positioned on the keyboard [Zhai, Hunter, and Smith, 2002]. This is particularly important for drivers who need to keep their eyes on the traffic. Drivers cannot afford to look for switches or control on the dashboard as this can lead to fatal accidents. Studies addressing one handed chording include [Gopher and Raij, 1985; Kirchenbaum, Friedman, and Melnik, 1986; Lyons, Plaisted, and Starner, 2004; Lyons, Starner et al., 2004; Noyes, 1983; Rosenberg and Slater, 1999; Seibel, 1962] and somehow related is also the half-QWERTY keyboard [Matias, MacKenzie, and Buxton, 1994] which is operated with one hand. Some studies also looked at two handed chording including [Beddoes and Hu, 1994; Gopher and Raij, 1985]. However, bimanual chording is impractical for in-car user interfaces since one hand is moved between the steering wheel and the gear stick.

Although chording allows fast text input, chording has a high learning threshold because the users must memorize the chord patterns. However, it is claimed that chording is easier to learn than QWERTY [Gopher and Raij, 1985]. The Microwriter was sold as a commercial device and employed a set of mixed mnemonic aids to simplify the learning and recall of chords. These memory aids included visual associations, word associations and kinaesthetic associations. Its documentation claims that users can learn to use the Microwriter in 2.5 hours. Training time for chording devices has also been greatly reduced for users capable of reading Braille [Cho et al., 2002; Sandnes and Huang, 2005, 2006a]. The 3x2 Braille patterns are used as two handed chords on bimanual chording devices with three keys for each
hand [Sandnes and Huang, 2006c]. Users are able to enter text with virtually no training as they can rely on their Braille knowledge. However, only few people are familiar with Braille. In this study a chording technique based on visual mnemonics first reported in [Sandnes, 2006] is presented. This strategy is easy to learn and robust to human error [Sandnes and Huang, 2006b]. The strategy will be outlined in the next section.

1.2 Chording with visual mnemonics

Mnemonics accelerate learning such as for memorizing virtual keyboard layouts [Lee and Zhai, 2004]. This study is based on visual mnemonics proposed in [Sandnes, 2006] to control in-car devices. This is illustrated in terms of text input. A character is entered in four steps. First, the user visualises the character as a 3x3 grid of pixels. Then, the user inputs this character grid in three steps from top to bottom, one scan-line at the time. For each chord one, two or three keys are used simultaneously. After the three chords are entered the required character is retrieved and the process is repeated.

For example, imagine the user wishes to input the character ‘A’. The user first visualises the character as a 3x3 pixel grid [see Fig. 1]. Then, the user enters the three chords – the first and the second scan-line requires the user to press all the three keys simultaneously (L+M+R) where L, M and R denote the left, middle and right keys, respectively, and then release the keys. The third chord comprises the left key (L) and the right key (R).

This strategy is similar to the iconic text entry strategy proposed by Jannotti [Jannotti, 2002], where characters are entered by pressing the keys on the 3x3 numeric keypads (digits 1 to 9) of a mobile phone in a shape that resembles the desired letter.

1.3 Chord memorization, recall and deduction

The input strategy assumes that the users are familiar with the visual appearance of the input symbols, and that they are able to mentally map these mental images onto the physical 3x3 pixel grid. The mappings for simple characters such as ‘T’ and ‘L’ can be achieved without training. More complex symbols such as ‘M’ and ‘G’ can be recalled successfully through a few training sessions. Users are able to memorise difficult patterns after seeing a few examples.

One advantage of the strategy discussed herein is that user do not have learn keyboard layouts or chord patterns since these can be inferred from the users own knowledge about the visual topology of the input symbols. The strategy can therefore be deployed with minimal amounts of training. Our implementation employs a one second timeout-reset that allows users to recover from mistakes or to start over after interruptions or pauses. After users learn how to abort the input of a character by pausing for one second they can correct simple errors without visual clues.
Figure 1: A 3x3 pixel grid is constructed by entering three three-finger chords in sequence from top to bottom. The technique is illustrated on a conventional QWERTY keyboard. The bottom of the illustration shows the chord matrix patterns for the English alphabet.

2 In-car user interface

2.1 Physical realization

The interaction style proposed herein can be realized as a chording keyboard attached to the steering wheel as shown in [Fig. 2] and [Fig. 3]. Drivers can then interact with the system without releasing the steering wheel. The keys should be physically placed in a position where the driver naturally holds the steering wheel while driving.

The chording alphabet used for the in-car user interface is a reduced version of that shown in [Fig. 1] and is illustrated in the subsequent sections.
2.2 Device mode selection

The in-car user interface illustrated in this paper controls a system comprising of a FM-radio, mp3-player, CD-player, DVD player, mobile phone, GPS navigation system and car lights. Each component is accessed by scrolling back and forward
through the list of devices using the symbols for \ (forward) and / (backward). The \ and the / symbols are easy to learn and are naturally associated with scrolling. The \ symbol is input by pressing the left, middle and right keys in rapid succession (see the character for backspace in [Fig. 1]). This symbol therefore mimics a left-to-right motion which matches Western user’s convention of what is next since Western users read from left to right [Sandnes, 2008].

Similarly, the / symbol is a mirror of \ and is input by the right, middle and left buttons in rapid successive keystrokes. This gives the impression of a right-to-left motion and hence the association to previous.

Device mode can also be selected directly by using a word association. By inputting the character for the first letter in the English word associated with the device the system enters the desired device state directly. Hence, R is used to represent Radio, C represents the CD-player, D represents the DVD-player, M represents the Mp3-player or Mobile phone (see section 3.7 for how ambiguities are resolved), T for mobile Telephone, N or G for GPS Navigation and L for lights.

<table>
<thead>
<tr>
<th>Modeless (volume)</th>
<th>Next/up</th>
<th>Previous/down</th>
<th>Toggle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M</td>
<td>1</td>
<td>M</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>M</td>
<td>3</td>
<td>M</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mode change</th>
<th>Next/up</th>
<th>Previous/down</th>
<th>Toggle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L</td>
<td>1</td>
<td>L</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>2</td>
<td>M</td>
</tr>
<tr>
<td>3</td>
<td>R</td>
<td>3</td>
<td>R</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Modal</th>
<th>Next/up</th>
<th>Previous/down</th>
<th>Toggle</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

**Figure 4: Modal and modeless next, previous and toggle commands.**

### 2.3 Modeless operations

Audio volume operations are applicable to multiple modes including the radio, mp3-player, CD-player and the DVD-player, and these are therefore assigned global modeless commands with universal meaning. Volume up is controlled by three repetitions of the chord middle+right keys [see Fig. 4] and volume down is controlled by the three repetitions of the chord left+middle keys [see Fig. 4]. Mute toggle is controlled by three repetitions of the chord left+middle+right [see Fig. 4].

### 2.4 Modal operations

An additional previous and next control is provided which meaning is dependent on the specific mode. Next is invoked by three repetitions of the left key (to symbolize a lefwards motion) and previous is invoked by three repetitions of the right key (to symbolize a rightwards motion).

Furthermore, a select toggle is indicated by three repetitions of the middle key. The modal chord sequences are illustrated in [Fig. 4].
The meaning of the modal selectors is listed in [Tab 1]. For the radio the next and previous are used to switch to the next or previous station, and the select toggle is used to switch between FM and AM bands. It is assumed that the radio is pre-tuned or capable of automatically tuning to the available stations. Manual tuning is therefore not needed.

For the CD-player and the mp3-player next and previous are used to move to the next and previous audio tracks and the toggle selector is used to pause and resume playback.

<table>
<thead>
<tr>
<th>Mode</th>
<th>Next →</th>
<th>Previous ←</th>
<th>Toggle</th>
</tr>
</thead>
<tbody>
<tr>
<td>R (Radio)</td>
<td>station</td>
<td>station</td>
<td>FM/AM</td>
</tr>
<tr>
<td>M (Mp3)</td>
<td>audio track</td>
<td>audio track</td>
<td>Pause/ resume</td>
</tr>
<tr>
<td>C (CD)</td>
<td>audio track</td>
<td>audio track</td>
<td>Pause/ resume</td>
</tr>
<tr>
<td>D (DVD)</td>
<td>chapter</td>
<td>chapter</td>
<td>Pause/ resume</td>
</tr>
<tr>
<td>T (Telephone)</td>
<td>Incomming call</td>
<td>contact</td>
<td>N/A Place call/ hang-up</td>
</tr>
<tr>
<td>G (GPS)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N (Navigation)</td>
<td>turn info</td>
<td>Turn info</td>
<td>Read current turn</td>
</tr>
<tr>
<td>L (Lights)</td>
<td>Next light</td>
<td>Previous light</td>
<td>Light on/ off</td>
</tr>
</tbody>
</table>

Table 1: The meaning of next, previous and toggle in the various device modes.

Next and previous in DVD-mode are used to move forward or backward one segment in the currently playing movie and toggle is used to pause and resume movie playback.

For the phone service next and previous are used to scan through the contact list and toggle is used to make and hang-up calls. If there is an incoming call the system will immediately go into mobile mode and the select toggle is used to accept the call.

In navigation mode next and previous are used to navigate to hear the next and the previous turning instructions and the select toggle is used to repeat the current turning instruction.

In car-light mode next and previous are used for cycling through the set of lights and toggle is used for switching the lights on or off.
2.5 Redundancy with mental icons

Some users are more visual than textual. A set of chord patterns was therefore based on a well-known set of A/V remote control icons [Diehl, 1994], i.e., play, pause, stop, fast forward and fast backwards. These icons are illustrated in [Fig. 5].

2.6 Audio feedback

Feedback is provided via speech whenever there is a change in system state, i.e., the system indicates to the user that the system has gone into a different mode, such as “Radio, playing channel 3 (followed by broadcast)”, or volume increased to 60% (followed by audio).

2.7 Error recovery and disambiguation

The chording strategy is robust to human error, and it has the ability to self correct input errors [Sandnes and Huang, 2006b]. This is especially true as only 25 of the 343 possible chord permutations are utilized (7%) which leaves plenty of redundancy.

Imagine that the three three-finger chord sequences associated with an input chord are represented by the bit vector $C$, and the set of symbols represented as bit vectors...
are $S$. Assuming that the Hamming distance between two bit vectors is given by $D(x,y)$, then the most probable input symbol is the one with the most similar bit pattern, namely

$$\arg\min_{s_j} D(c,s_j) \forall s_j \in S$$

(1)

There is also an ambiguity for the input character $M$ which is assigned to both mobile phone and the mp3 player. If the user is in mobile phone mode, then the $M$ symbol is interpreted as mp3-mode change. If however, the system is in mp3 mode the $M$ is interpreted as a change into mobile phone mode.

2.8 System state and memory

The system stores the state of each mode as the user changes the modes. This allows users to resume where they left off once they return to a mode. For instance, if the user is listening to the radio and there is an incoming phone call and the system automatically goes into incoming phone call mode and the radio is muted. Once the phone call is finished the system returns to the radio mode with the same radio station and volume. Furthermore, persistent mode state is also useful if the user accidentally move to a different state than the one intended.

3 Experimental evaluation

To test the feasibility of the interaction strategy a text input experiment was conducted, which allowed us to measure how easy it is to learn to use the system, users’ performance and to verify if the strategy is reliant on visual feedback or not.

3.1 Participants

A total of 14 participants were recruited for the experiment. Subjects were recruited from the student population of Oslo University College. They were all male computer science students aging from 21 to 38 years with a mean age of 26 years. All the participants reported being right-handed, having no visual impairment or dyslexia. Each participant was paid 100 Norwegian kroner for each hour of participation in the experiment. A bonus of 100 kroner was awarded upon completing all the sessions.

3.2 Experimental setup

Desktop computers with monitors and full QWERTY keyboards were used to ensure a repeatable and controlled experiment. The software was custom implemented as two Java Applets running in a web browser. One applet comprises the character pattern reference and is presented as a virtual QWERTY keyboard. By clicking on one of its keys a graphical representation of the corresponding 3x3 pixel map for the character is displayed for the duration of 1 second. The second applet comprises the text entry application. The applet has two components. The top line shows a phrase to be copied and at the bottom a multi-line text entry area is provided.

As users enter text the text is displayed in the text entry area which behaves in the same manner as a text-area widget. No intermediate state information is shown on
the screen during the assembly of a character – this allows the claim of character
construction with low visual attention to be demonstrated. The text scrolls
automatically so that it is possible to see a trail of previously entered phrases. The
right side of the screen was used to display short instructions on how to operate the
system. Both applets logged user activity locally to disk.

The applet was configured to accept keystrokes from the SPACE, J and K keys
as this is a hand position touch typists are used to that naturally fits the thumb, index
finger and long finger of the right hand. This is an asymmetric finger configuration
that may introduce some stimulus-response incompatibility as the SPACE-bar is not
aligned on the same line as J and K. I.e., the finger positions are two-dimensional
while the chord scan-line is one-dimensional. However, experimentation revealed that
this is more comfortable than using the three keys in a row such as H, J and K.

3.3 Procedure

The experiment was carried out in isolation. Each participant was instructed in how to
operate the system and the text entry task to be conducted. The principal investigator
was present during the text entry sessions to answer questions and provide help. Each
session lasted one hour and a total of four sessions were carried out for each of the 14
participants. Each session was separated by approximately three days.

3.4 Materials

Four different lists of phrases randomly selected from the phrase sets described in
[MacKenzie and Soukoreff, 2003] were used in the experiments. Each phrase list, of
approximately equal difficulty, was used for each of the four sessions to eliminate
phrase learning effects. All phrases were presented in lowercase as lowercase
characters are easier to read, and the phrases did not contain special characters or
punctuation symbols. The users had to perform an implicit mental transformation
from lower to upper case as the text was mostly displayed as lower case characters,
while most characters are input as uppercase characters.

3.5 Task

The participants were asked to conduct a text copying task where the phrases were
displayed one by one. Once completing a phrase the user had to enter the chord
sequences for ENTER to display the next phrase. There was no facility for going back
to the previous phrase. Subjects were instructed to correct errors using the chord
sequences representing BACKSPACE.

3.6 Measurements

The character reference applet logged the character looked up by the subjects and the
time at which the reference was made. The text entry applet logged all the keystrokes
with associated timestamps and the input phrases and text output by the system.
3.7 Analysis

Relevant information was extracted from the log files using scripts. The various parameters discussed herein are based on means for each subject over an entire session.

Statistical significance tests were performed using both a one-way ANOVA and two-way ANOVA using the Microsoft Excel analysis tool-pack with a significance level of 0.05. The session was the independent variable, the subjects were used as replications (no between subject analysis was performed) and the dependent variables include error-rates and text entry speeds.

![Figure 6: Mean text entry performance in words per minute (WPM). Error-bars show standard deviation.](image)

4 Results

4.1 Performance

[Fig. 6] shows the mean words per minute (WPM) based on the means for the fourteen subjects plotted across the four sessions. Note that these statistics only include text that is correctly input, and incorrectly entered text and other characters, including BACKSPACE and ENTER, are discarded. Furthermore, the time delay that occurs between displaying a new phrase and the first keystroke is also discarded to minimise reading effects. The plot reveals that there is a steady increase in productivity from a mean of 2.9 WPM during the practice session to 5.9 WPM during the fourth session. The effect of practice is statistically significant (F(3,48)=6.06; p<0.001). The slowest participant enters text at a mean rate of 2.4 WPM and the fastest subject enters text at a mean rate of 9.9 WPM during the fourth session. The results demonstrate that the participants gradually improve their text input skills when using the technique.

The text entry speeds measured do not match those of traditional chording such as the chording glove [Rosenberg and Slater, 1999] where a mean text entry rates of 8.9 WPM was achieved during the practice session and 12.2 WPM during the fourth practice session. This is not surprising since three chords must be entered in succession to retrieve a character instead of just one chord with the chording glove. However, the strategy studied herein based on spatial mnemonics has a steeper
learning rate. In [Fig. 7] the performance curve of the current approach is plotted in a log diagram together with the performance curve reported for the chording glove, where the number of practice trials is represented by the number of hours of using the system, and the response time is represented by the text entry rate. Although the chording glove yields a higher text input rate (a high $A$ value) the slope of the curve is smaller ($B = 0.21$) than that for the current strategy ($B = 0.27$). This comparison shows that the learning rate for participants using the spatial mnemonics is higher than when using traditional chords.

Figure 7: The power law of practice for the spatial mnemonics approach versus traditional chords (chording glove).

4.2 Error

The mean error rate for all the four sessions is 10.7% with a large spread (SD = 7.4) and no significant effect on practice ($F(1,52)=0.41;p>0.75$). [Fig. 8] illustrates how the errors are distributed at chord level. The participants made the fewest mistakes entering the first chord and the most mistakes entering the second chord ($F(2,156)=34.9;p<0.001$).

It was suspected that the high overall error rate of 10.7% was partially caused by trial and error behaviour, as most literature on text entry report error rates of about 5%. To test this hypothesis trial and error sub-sequences in the entered text were identified, namely ch + BACKSPACE + ch + BACKSPACE + ... + ch + BACKSPACE, where ch represents arbitrary characters. It is assumed that the participants first enters a character, is not satisfied, enters backspace and tries again. The analysis show that 33.7% (SD=14.7) of all self-reported (backspace) errors are...
associated with trial and error sequences. This large ratio confirms the hypothesis that trial and error behaviour accounts for a large chunk of the overall error rate. The mean portion of trial-and-error sequences decreased non-significantly with practice (F(3,52)=1.7;p>0.17). [Fig. 9] shows the mean number of attempts needed to retrieve the desired characters. During the practice session a mean of 3.8 attempts were made before the desired character was retrieved and this mean falls to 2.7 attempts for the final session. Therefore, practice helps reduce the number of attempts (F(1,52)=48.8;p<0.001). Since the mean number of attempts decreases with practice, and the ratio of trial and error does not, the participants use trial and error more often when gaining practice.

Figure 8: Distribution of chord-level errors (percentage of observed errors). Error-bars show standard deviation.

Figure 9: Mean number of trial-and-errors per character. Error-bars show standard deviation.
Learning

In order to assess learning the number of accesses the participants made to the character reference during each text entry session was analysed. [Fig. 10] shows the mean learning curve for the 14 participants across the four sessions. The vertical axis represents the number of accesses made to the character reference table. The results show that the number of references to the character reference manual decreases for each session (F(3,52)=7.03; p<0.001).

During the third session three of the participants requested help less than 10 times. During the fourth session nine participants requested help less than 10 times of which two participants required no help at all. This indicates learning, and the results therefore confirm that it is possible to memorise or infer the patterns based on spatial mnemonics with four hours of practice. The results are consistent for all the participants with the exception of one participant who required noticeably more help than the others. However, for this participant there is also a distinct learning curve, i.e., the slowest learner requested help 397 times during the practice session and only 149 times during the final session.

The reference material was consulted less in the current study than for the chording glove [Rosenberg and Slater, 1999], where more than 1% of the time was spent consulting the reference material during the fourth hour of practice, and this fell to 0.4% during the eight session. Note however that the details regarding how these measurements were conducted are sketchy.

Certain characters are more difficult to recall than others (F(25,1352)=3.1;p<0.001). Further, practice has a significant effect on learning (F(3,1352)=51.4;p<0.001).

5 Limitations of this study

This study is based on a small and relatively homogenous sample of young adults. Related studies show that results can differ significantly across different user groups, for example young users versus old users, or females versus males. One would expect to find that young adults would perform significantly better than old users. Further, the spatial nature of the text input approach suggests that there could possibly be a
gender dependent performance difference. Note that these tests were conducted as a primary task and interferences between simultaneous tasks were not measured.

6 Summary

An in-car user interface interaction style was proposed based on three finger chord sequences. The chords can be input with limited visual feedback and are memorized with little training. An in-car menu control system based on the proposed interaction style was outlined. Experimental results confirmed that the input strategy can be learned quickly, provides adequate performance and requires close to no visual attention.

References


