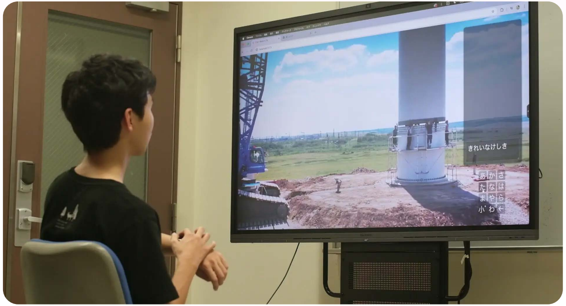
Flick-in

ACM SUI 2025, 11 November 2025

Japanese Text Entry Method for Indirect Touch Using Bezel-initiated Swipe

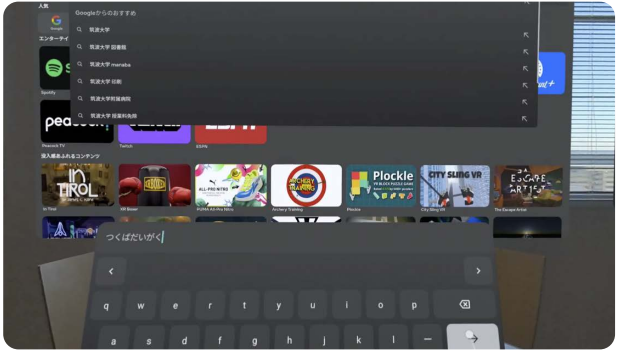


Yuto Wada, Myungguen Choi, Kaoru Shirane, Buntarou Shizuki
University of Tsukuba / wada@iplab.cs.tsukuba.ac.jp

Introduction

Text entry on smart TV and MR environment

e.g.
- Post comments while watching videos
- Send messages while looking at another user’s avatar



Introduction

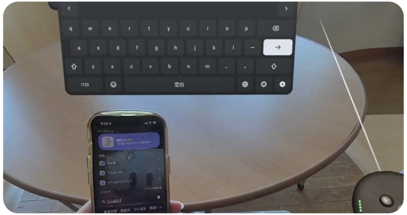
Pointing at on-screen keyboard

Large keyboard obstruct the underlying content.

Voice input

Difficult to use in noisy environments and public spaces [1].

External devices
(e.g. smartphone)

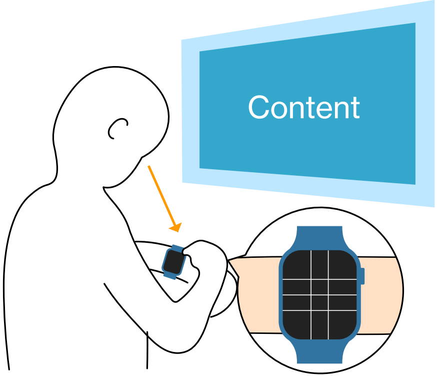


Users need to shift their visual attention between the device and the display.

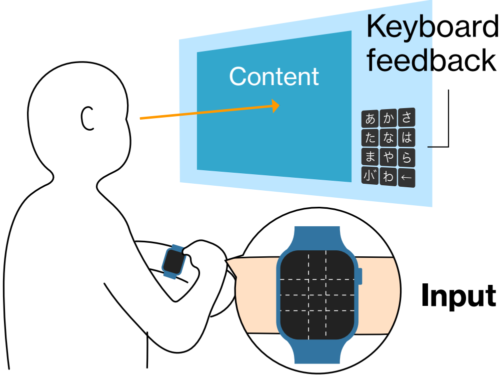
Conventional text entry methods have several limitations

[1] S. Ronkainen, et al. Tap Input as an Embedded Interaction Method for Mobile Devices. TEI ‘07, pp. 263–270, 2007.

Conventional Method

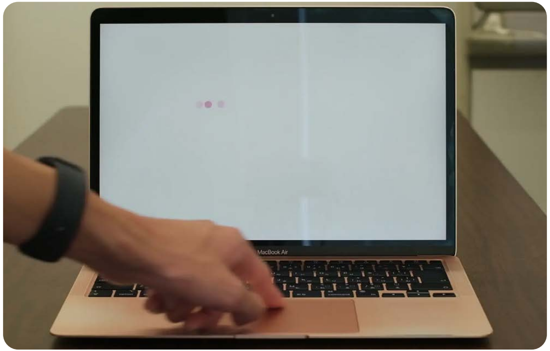


Indirect touch typing



Indirect touch

Touch interactions where an input surface and an output surface are separated [2,3].



Challenge

Touching down at the precise location is difficult, since users do not look at the input surface [4].

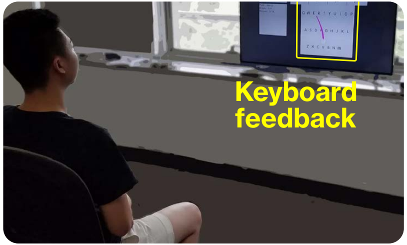
[2] Z. Yang, et al. Investigating Gesture Typing for Indirect Touch. IMWUT, Vol. 3, pp. 117:1–117:22, 2019.
[3] S. Voelker, et al. Combining Direct and Indirect Touch Input for Interactive Workspaces using Gaze Input. SUI '15, pp. 79–88, 2015.
[4] Y. Xin, et al. Strategy for Improving Target Selection Accuracy in Indirect Touch Input. IEICE Transactions on Information and Systems Vol. E103, Issue 7, pp. 1703–1709, 2020.

Introduction

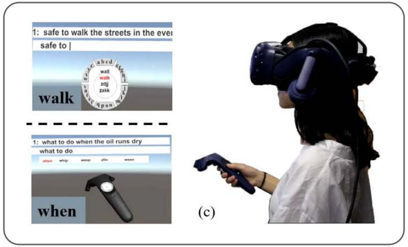
Indirect touch typing

- Allowing for accurate text entry by relying on **precise touch-up gestures**
- Displaying a keyboard feedback and a pointer on the output surface
- Slide gesture to move the pointer ➡ Touch-up gesture to select a key

G-keyboard [1]



HiPad [5]



HoldBoard [6]

[5] H. Jiang, et al. HiPad: Text entry for Head-Mounted Displays Using Circular Touchpad. IEEE VR 2020, pp. 692–703, 2020.
[6] S. Ahn, et al. Typing on a Smartwatch for Smart Glasses. ISS '17, pp. 201–209, 2017.

Introduction

Most method are primarily designed for English using QWERTY keyboards.

There is a lack of research on methods for Japanese.

[5] H. Jiang, et al. HiPad: Text entry for Head-Mounted Displays Using Circular Touchpad. IEEE VR 2020, pp. 692–703, 2020.
[6] S. Ahn, et al. Typing on a Smartwatch for Smart Glasses. ISS '17, pp. 201–209, 2017.

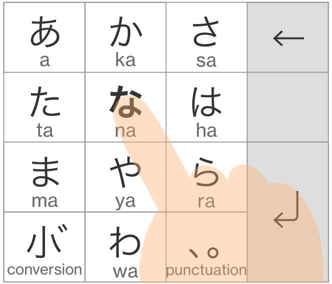
Kana letters

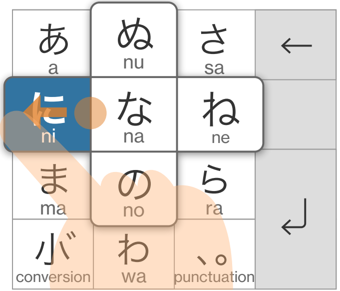
- Japanese text is fundamentally based on 50 Japanese syllabary letters (kana letters).
- Transcribed into a combination of **10 basic consonants** (‘a’, ‘k’, ‘s’, ‘t’, ‘n’, ‘h’, ‘m’, ‘y’, ‘r’, ‘w’) and **5 vowels** (‘a’, ‘i’, ‘u’, ‘e’, ‘o’).

Vowel	Basic consonant: あ a	か k	さ s	た t	な n	は h	ま m	や y	ら r	わ w
あ a	あ a	か ka	さ sa	た ta	な na	は ha	ま ma	や ya	ら ra	わ wa
い i	い i	き ki	し shi	ち chi	に ni	ひ hi	み mi		り ri	を wo
う u	う u	く ku	す su	つ tsu	ぬ nu	ふ hu	む mu	ゆ yu	る ru	ん n
え e	え e	け ke	せ se	て te	ね ne	へ he	め me		れ re	ー
お o	お o	こ ko	そ so	と to	の no	ほ ho	も mo	よ yo	ろ ro	

Flick method

Widely used on devices with a touchscreen





1. Consonant selection

Touch down on the consonant key

2. Vowel selection

Touch up (‘a’) or Flick in the corresponding direction (‘i,’ ‘u,’ ‘e,’ ‘o’)

Flick-in

Improving the Flick method for indirect touch on a smartwatch

Design

1. Vowel Selection

Users select a vowel using either
a touch-down gesture or BIS (Bezel-initiated swipe).

Design

1. Vowel Selection

When the target vowel is ‘a,’
users perform a touch-down gesture near the center.

Design

1. Vowel Selection

When the target vowel is ‘i,’ ‘u,’ ‘e,’ or ‘o,’
users slide their finger from the corresponding bezel into the touchscreen.

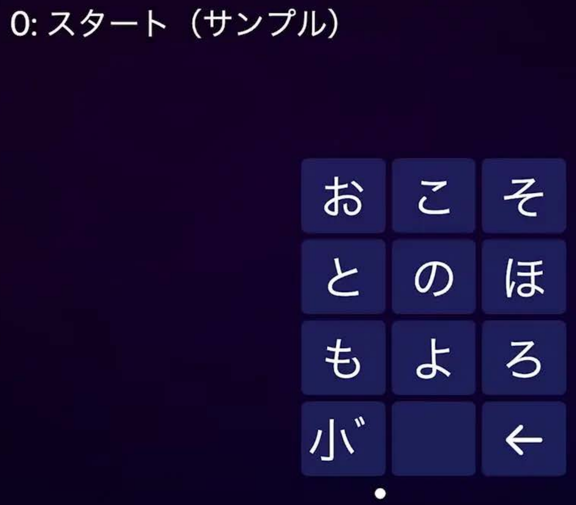
Design

2. Consonant Selection

After a vowel is selected, the keyboard feedback updates to display the keys corresponding to the kana letters with the selected vowel.

Design

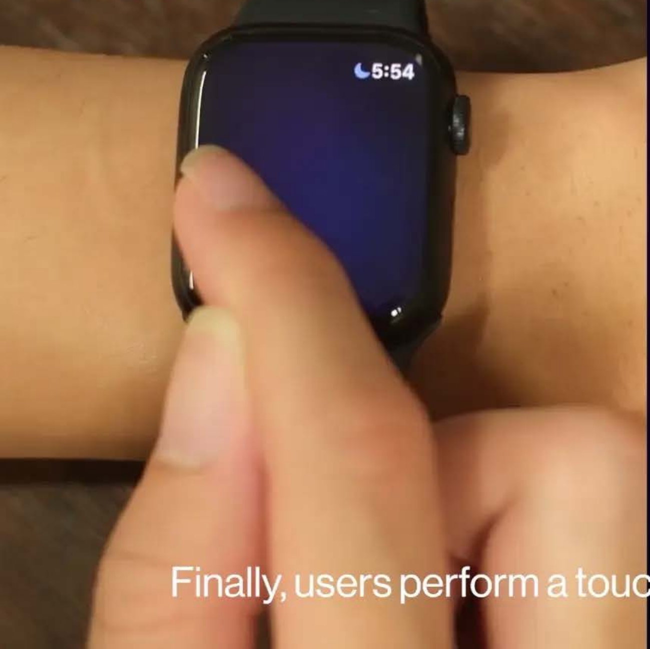
2. Consonant Selection



Users slide their finger to move the pointer to the target consonant key while observing the keyboard feedback and the pointer.

Design

2. Consonant Selection



0: スタート（サンプル）

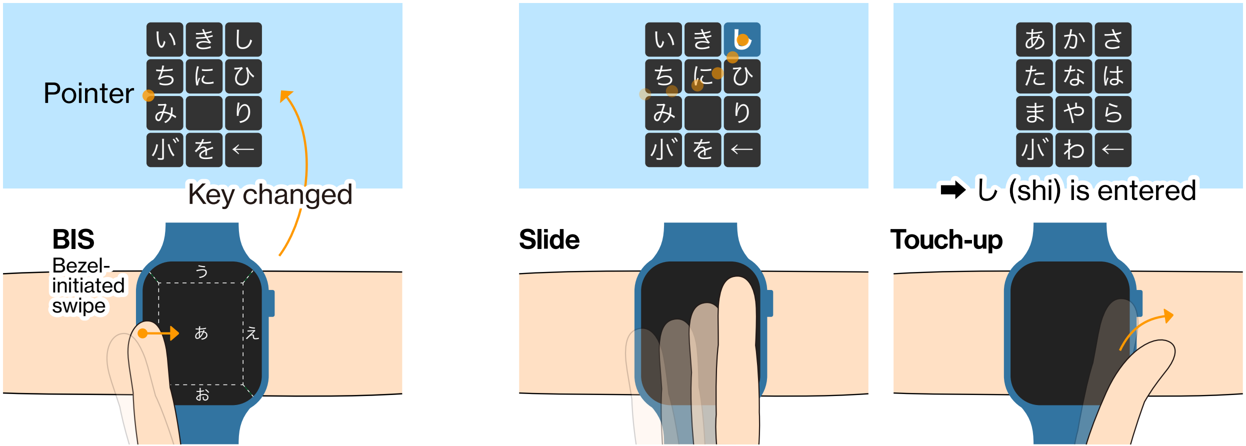
Finally, users perform a touch-up gesture to select the consonant.

Flick-in – Design

Reversing the input order of vowels and consonants compared to the Flick method

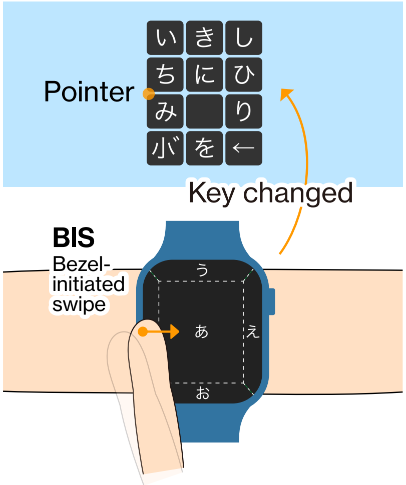
1. Vowel Selection

2. Consonant Selection



Flick-in – Design

1. Vowel Selection



Bezel-initiated Swipe (BIS)

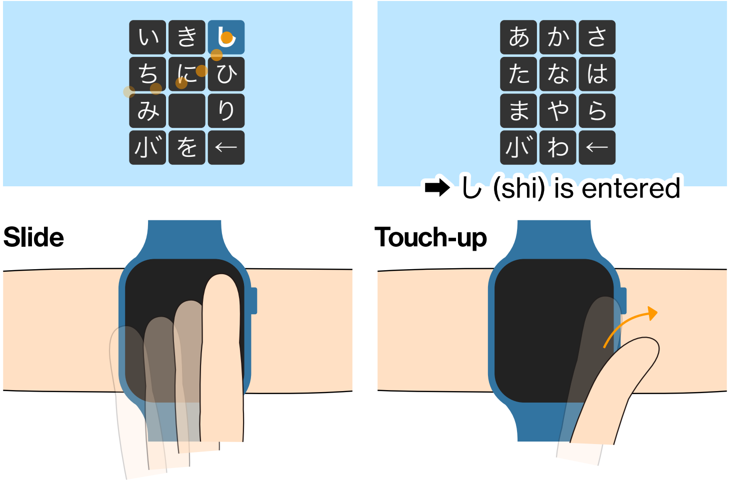
BIS with fewer input directions [7] and touch-down gestures can be performed roughly.

Flick-in achieves accurate vowel selection by assigning the small number of vowels to either the BIS or touch-down gesture.

[7] P. C. Wong, et al. Exploring Eyes-free Bezel-initiated Swipe on Round Smartwatches. CHI '20, pp. 1–11, 2020.

Flick-in – Design

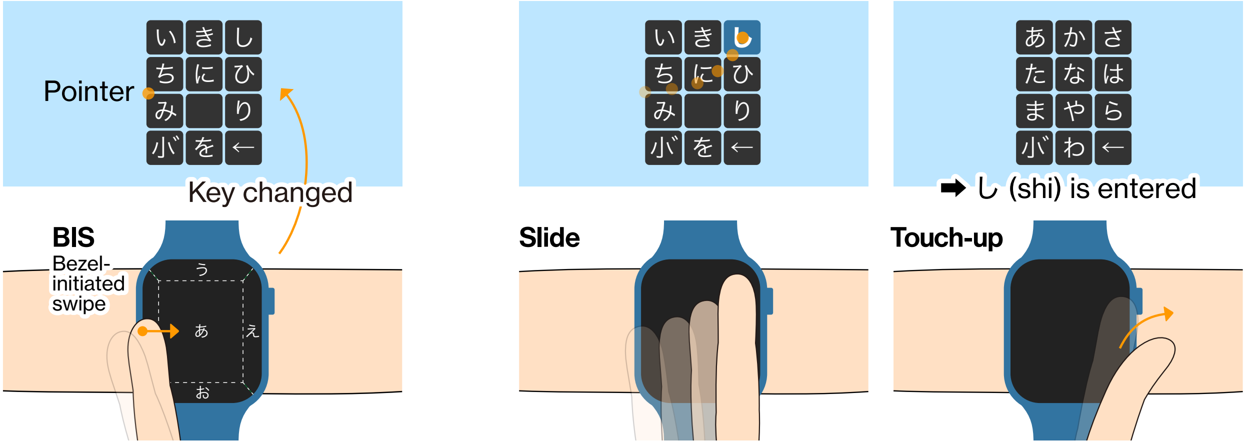
2. Consonant Selection



Pointer operation with slide gesture & key selection with touch-up gesture

Although the number of consonants is large, users can accurately select keys via a pointer.

Flick-in – Design

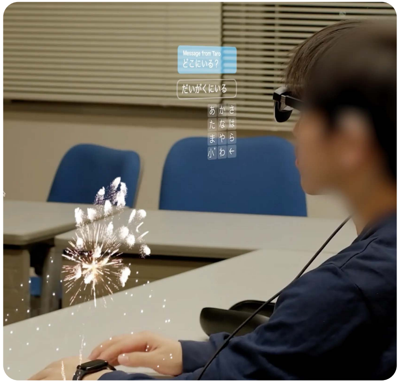


➡ User can select vowels and consonants accurately without visually confirming the input surface.

Flick-in – Characteristics

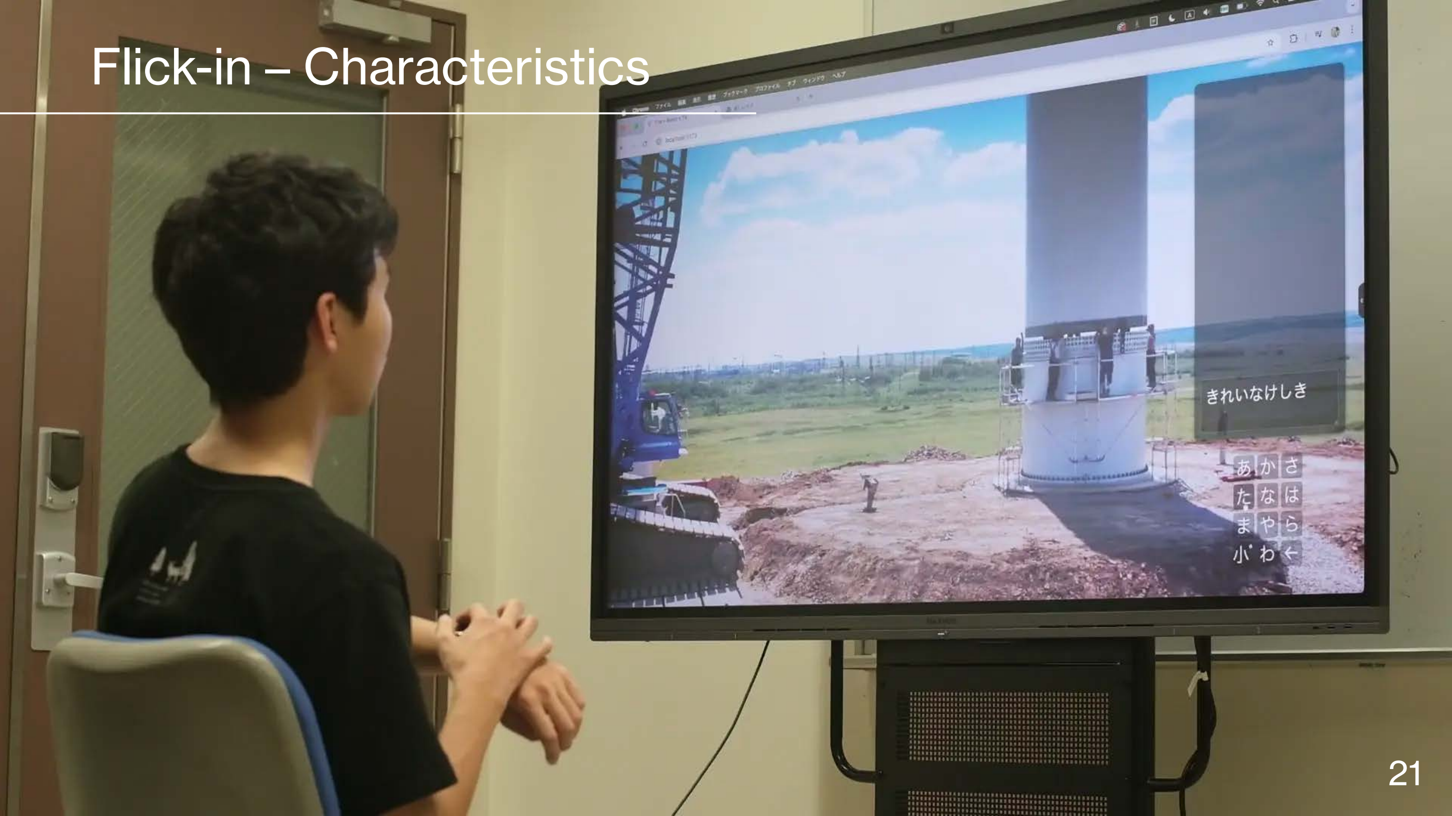
Usable with
indirect touch

Usable in
various postures
(e.g. walking)

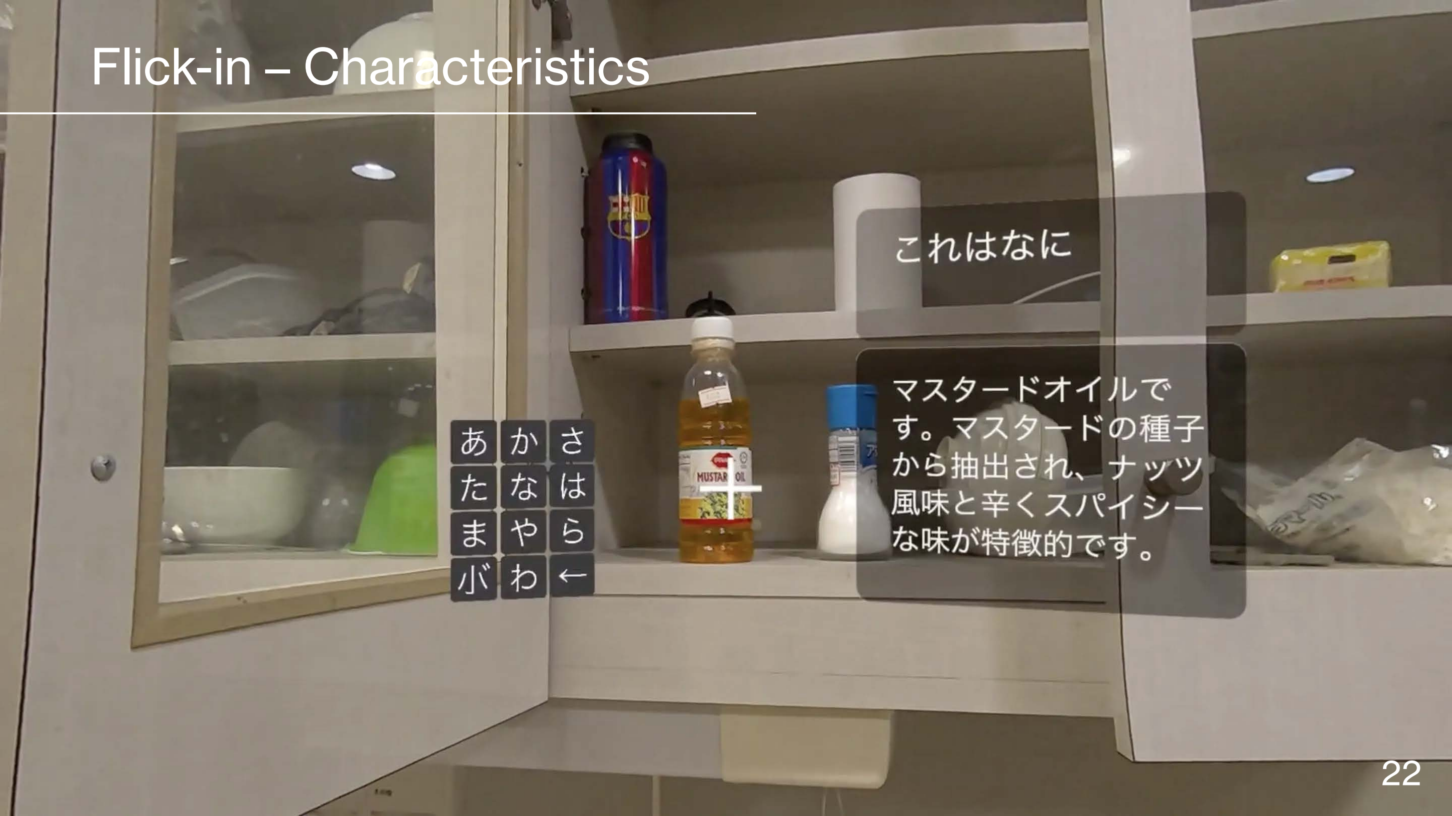


Usable with
small keyboard feedback

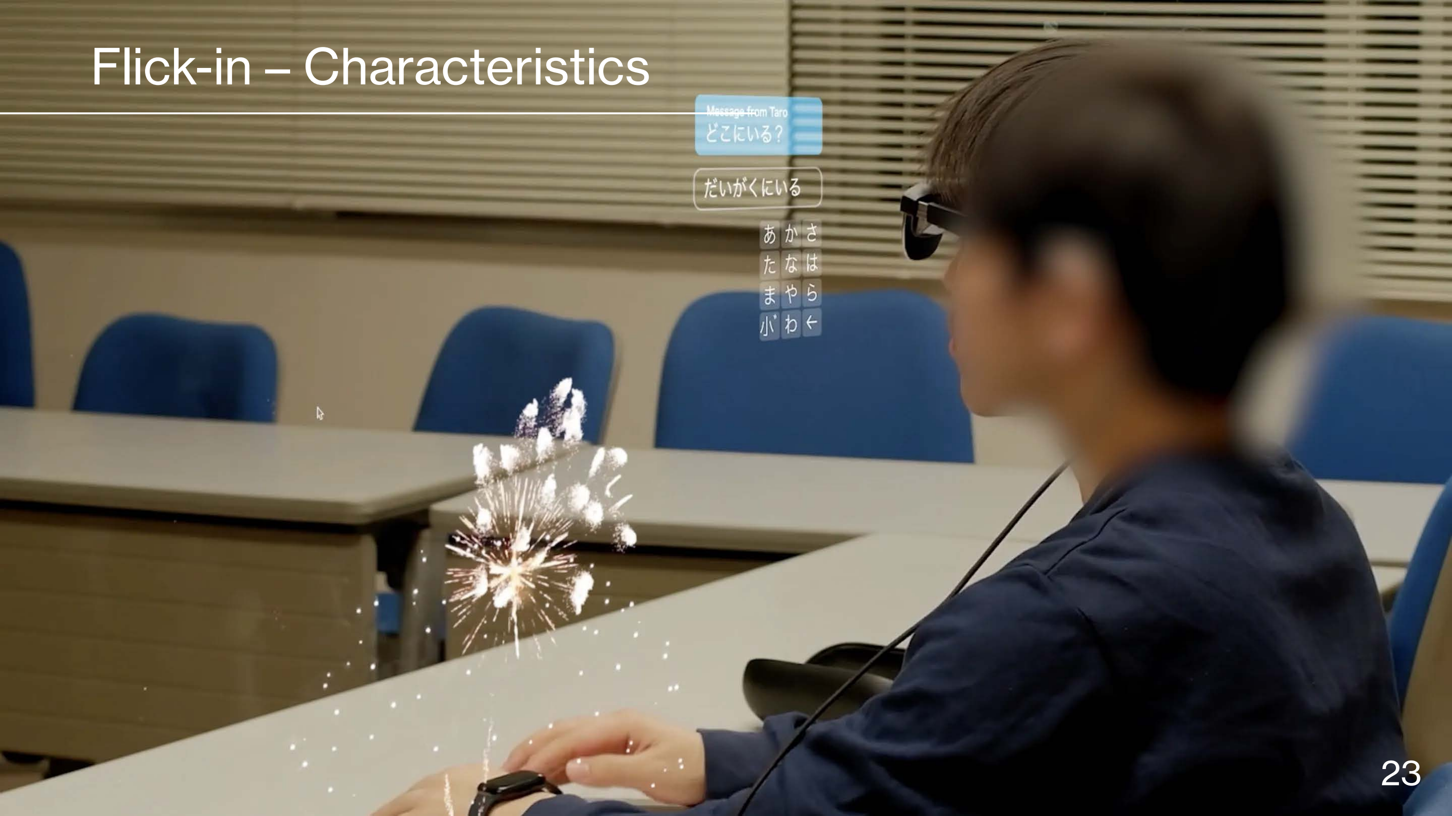
Flick-in – Characteristics



Flick-in – Characteristics



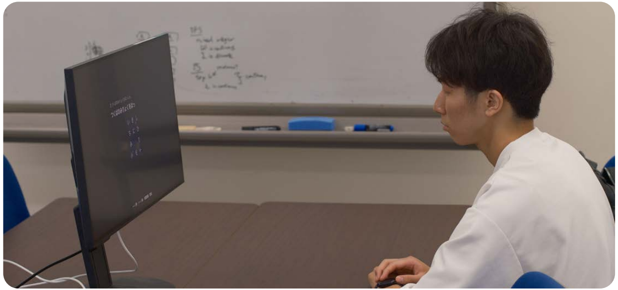
Flick-in – Characteristics



User Study

We conducted the user study for **the text entry task** twice to evaluate the performance of Flick-in.

User Study 1
with an external display

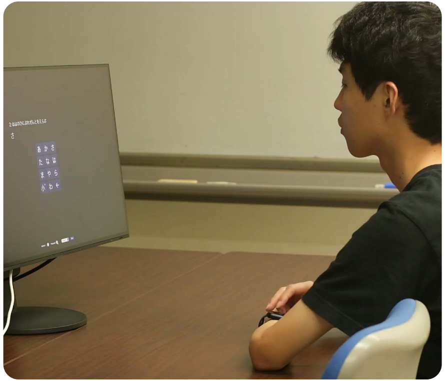


User Study 2
in an MR environment

User Study 1 – Overview

Purpose

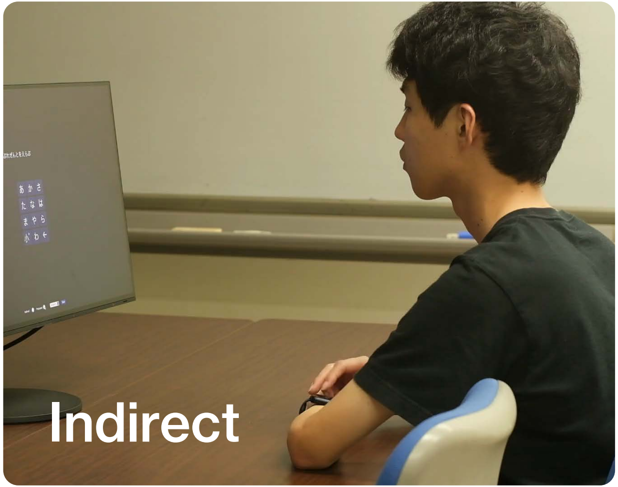
Performance evaluation of Flick-in in indirect/direct touch through comparison with Flick

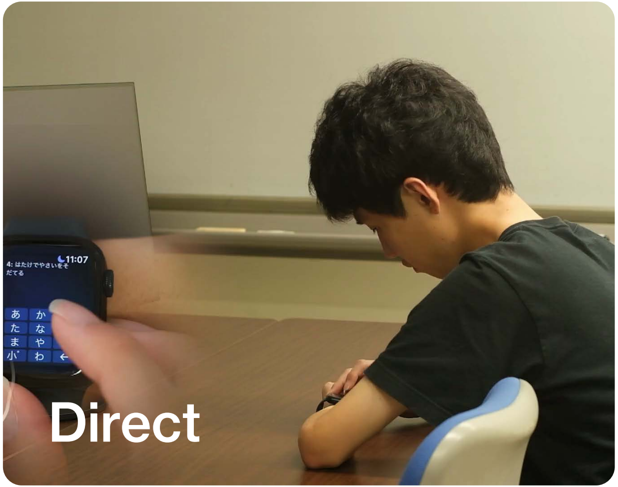


- Participants transcribe the sentences as quickly and accurately as possible.
- Within-subjects design (12 participants, mean age=22.8 years)
- Metrics: Character Per Minutes (CPM), Total Error Rate, System Usability Scale, NASA-Raw Task Load Index

User Study 1 – Independent variables (1/2)

Touch


Indirect



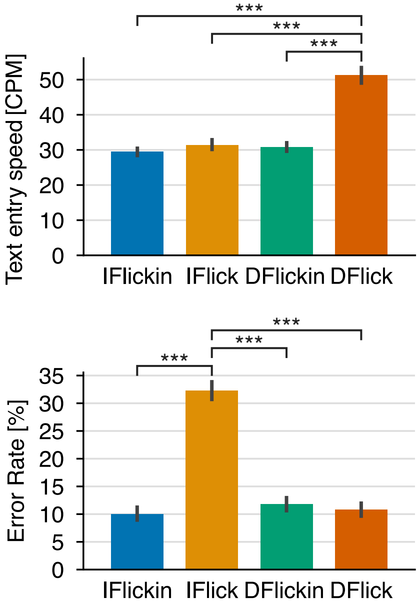
Direct

Method


11:07
3: はるはわかれのきせつ
はるはわかれの
Flickin
Our method


11:07
4: はたけでやさいをそだてる
はたけ
Flick
Baseline method

User Study 1 – Text entry speed and error rate

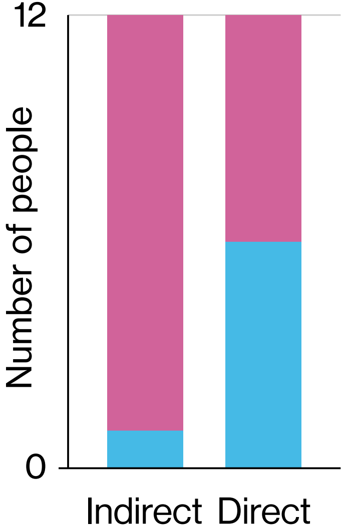


	CPM	TER
Indirect-Flickin	29.5 CPM	10.0%
Indirect-Flick	31.4 CPM	32.3%
Direct-Flickin	30.8 CPM	11.8%
Direct-Flick	51.2 CPM	10.8%

In indirect touch

Flick-in has a significantly lower error rate, while having a comparable speed to Flick.

User Study 1 – Preferences



Indirect touch

Flick-in **11 participants**

"Lower error rate"

"Ability to cancel input"

Flick **1 participant**

"Used to use"

Direct touch

Flick-in **6 participants**

"Easy selection of the small keys"

Flick **6 participants**

"Method they are familiar with"

User Study 2 – Overview

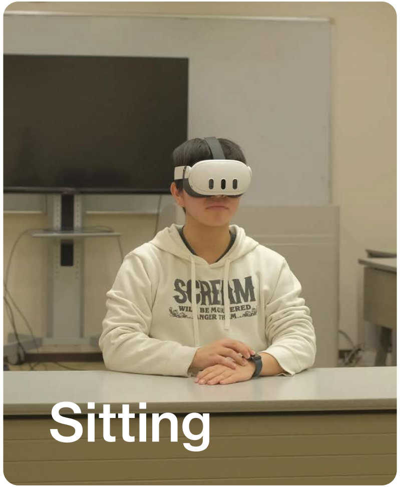
Purpose

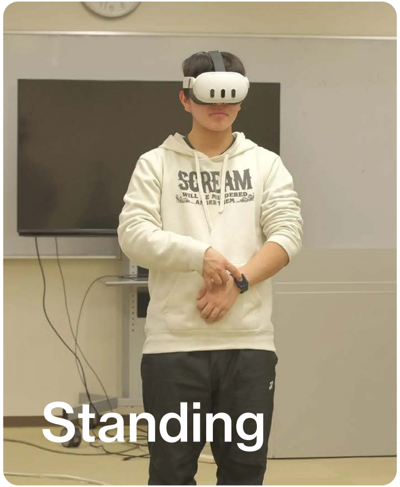
Performance evaluation of Flick-in in an MR environment

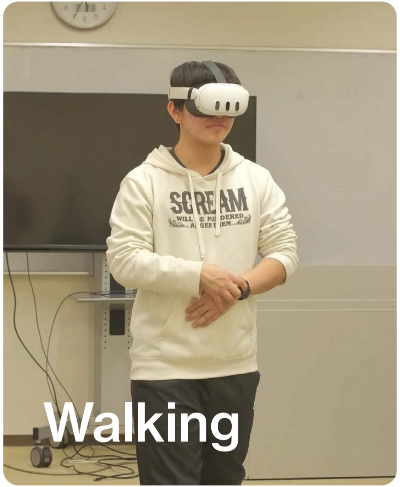
- The task and metrics are the same as those of Study 1.
- Within-subjects design (12 participants, mean age=22.3 years)

User Study 2 – Independent variables (1/2)

Posture







User Study 2 – Independent variables (2/2)

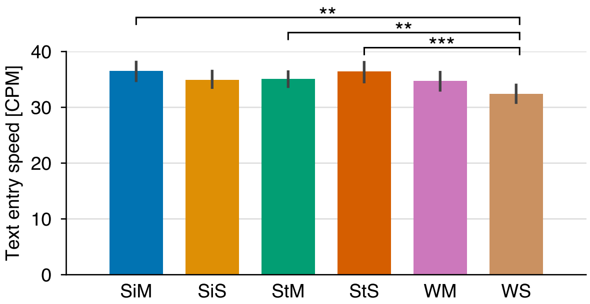
Size Keyboard feedback size

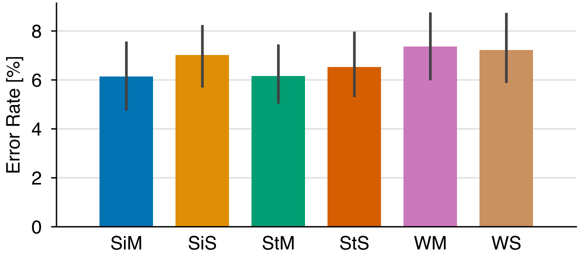
Medium (64 dmm)

Small (24 dmm)

dmm (distance-independent millimeter): A size that appears 1 mm at a distance of 1 m

User Study 2 – Text entry speed and error rate





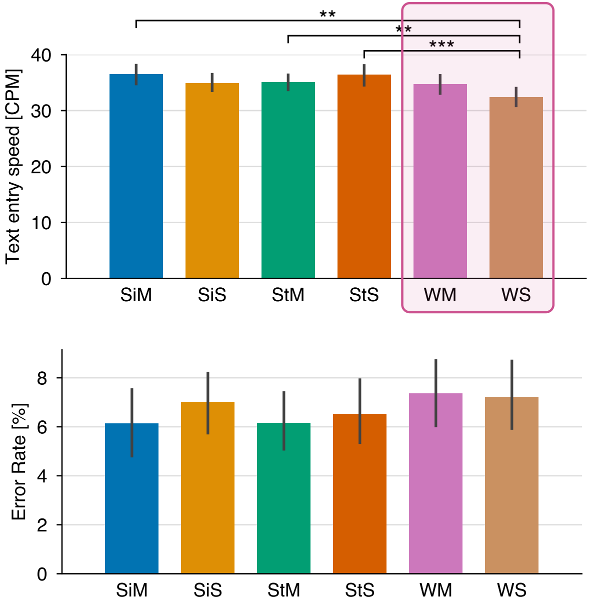
		CPM	TER
Sitting-Medium	(SiM)	36.5 CPM	6.14%
Sitting-Small	(SiS)	34.9 CPM	7.02%
Standing-Medium	(StM)	35.1 CPM	6.16%
Standing-Small	(StS)	36.4 CPM	6.53%
Walking-Medium	(WM)	34.7 CPM	7.37%
Walking-Small	(WS)	32.4 CPM	7.23%

Posture

Walking is significantly slower than the other postures, but the decrease is small and the speed remains practical.

Size No significant difference

User Study 2 – Text entry speed and error rate



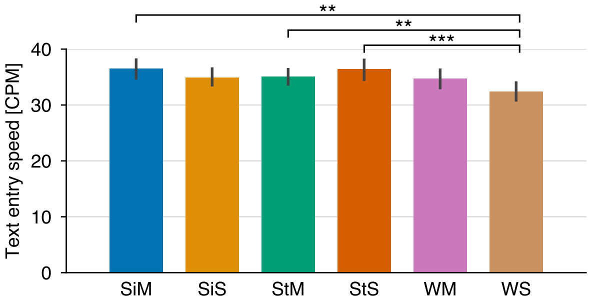
		CPM	TER
Sitting-Medium	(SiM)	36.5 CPM	6.14%
Sitting-Small	(SiS)	34.9 CPM	7.02%
Standing-Medium	(StM)	35.1 CPM	6.16%
Standing-Small	(StS)	36.4 CPM	6.53%
Walking-Medium	(WM)	34.7 CPM	7.37%
Walking-Small	(WS)	32.4 CPM	7.23%

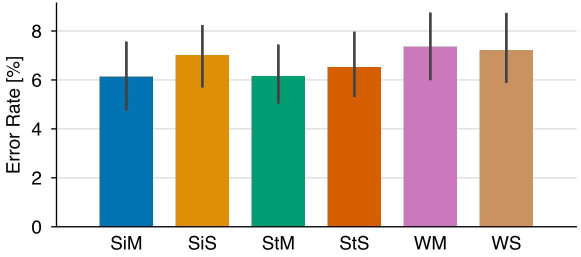
Posture

Walking is significantly slower than the other postures, but the decrease is small and the speed remains practical.

Size No significant difference

User Study 2 – Text entry speed and error rate





		CPM	TER
Sitting-Medium	(SiM)	36.5 CPM	6.14%
Sitting-Small	(SiS)	34.9 CPM	7.02%
Standing-Medium	(StM)	35.1 CPM	6.16%
Standing-Small	(StS)	36.4 CPM	6.53%
Walking-Medium	(WM)	34.7 CPM	7.37%
Walking-Small	(WS)	32.4 CPM	7.23%

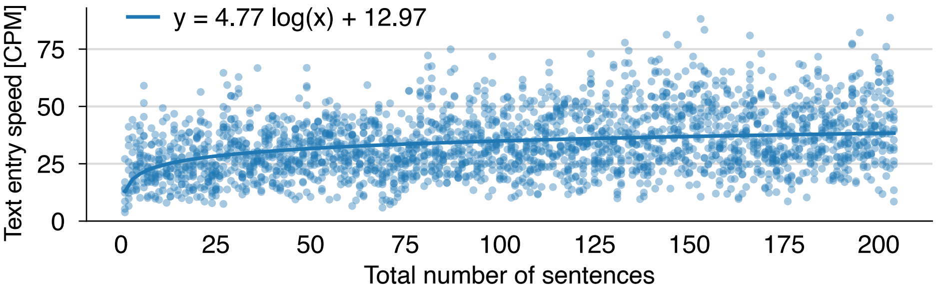
Posture

Walking is significantly slower than the other postures, but the decrease is small and the speed remains practical.

Size No significant difference

Discussion – Proficiency

Study 2 Although the input order of vowels and consonants is oppposited to the baseline method, Flick-in achieved approximately 40 CPM after about 200 inputs.



➡ CPM could improve as users become more proficient with Flick-in.

Discussion – Future works

Long-term study

To better understand its performance in daily use.

Kana-kanji conversion function

Flick-in can incorporate a conversion function by displaying conversion candidates.

Kana　しこう

↓ **Conversion**

Kanji　思考
thinking

至高
supreme

Optimization of design

Some participants mentioned a sense of discomfort with the direction of vowel selection.

Flick-in Japanese text entry method for indirect touch

In Flick-in, vowel input is assigned to BIS and performed in advance.

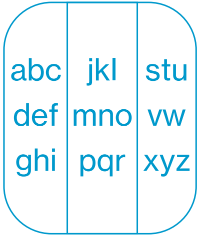
- **Study 1**
  Flick-in achieves text entry with significantly higher accuracy than Flick in indirect touch scenario.
- **Study 2**
  Flick-in maintains stable typing across various postures even with small keyboard feedback size.

Appendix: Application to other languages

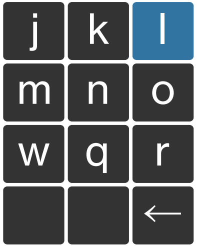
Flick-in could be adapted for other languages if their characters can be classified into groups.

e.g. English

1. BIS or touch-down
Selecting analphabet group, devided by following previous studies [8,9]



2. Pointer selection
Choosing a character within the group.

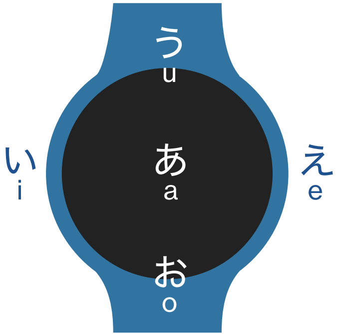


➡ Allowing for larger key sizes than those of the QWERTY keyboards

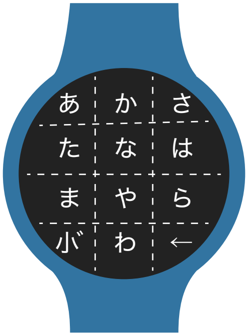
[8] S. Ahn, et al. Gaze-Assisted Typing for Smart Glasses. UIST '19, pp. 857–869, 2019.
[9] XA Chen, et al. Swipeboard: a Text Entry Technique for Ultra-small Interfaces that Supports Novice to Expert Transitions. UIST '14, pp. 615–620, 2014.

Appendix: Using other devices

BIS on a circular smartwatches
can be performed accurately (93.34% in six directions) [7].



Vowel selection



Consonant selection

Touchpad on
an handheld controller [10]

[10] HTC. Buy VIVE Hardware. https://www.vive.com/sea/product/vive/ (Access: 2025-11-11)

Appendix: Similarities with the Flick method

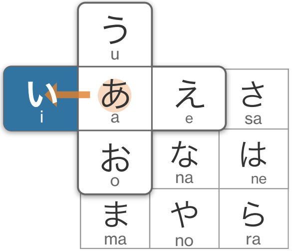
Vowel

Commonalizing the starting point of BIS with the flick direction

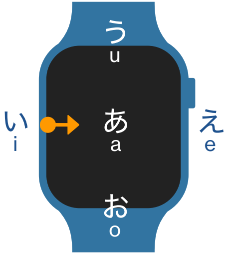
Consonant

Adopting the keyboard of the Flick method

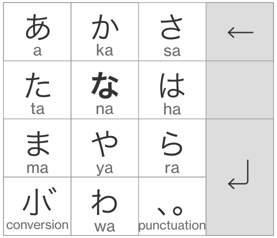
Flick



Flick-in



Flick

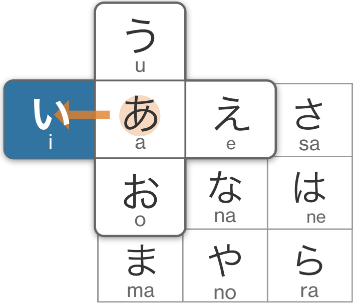


Flick-in

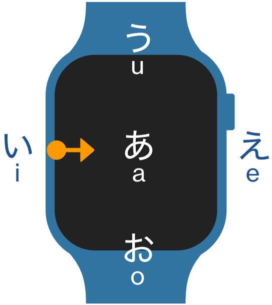
Appendix: BIS direction

The BIS direction in Flick-in is opposite to the flick direction of the Flick method.

➡ Some participants reported discomfort.



Flick: Right to left



Flick-in: Left to right

Appendix: Design Q&A

Why the input order is reversed compared to the Flick method?

- To achieve accurate input.
- Japanese: 10 consonant (relatively many) + 5 vowels (few)
- Selecting a vowel first enables user to initiate a gesture roughly.

Why does the Flick-in use both touch-down and BIS gestures?

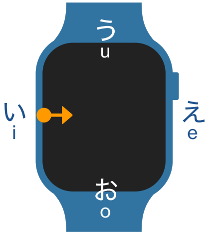
Touch-down

1 vowel (a)

+ **BIS**

4 vowels

(i, u, e, o)



= 5 vowels

Appendix: Design Q&A

Is it difficult to become proficient of Flick-in since the input order is reversed?

Some participants were initially a bit confused
but they became accustomed to using it over time.

What are the advantages of Flick-in?

Although Flick-in does not improve in speed,
the advantage of Flick is its reduction of error rates.

The error rate under the indirect condition

Flick-in: 10.0% —Significant higher→ Flick: 32.3%

Appendix: User Study – Procedure

1. **Study description**
2. **Pre-study questionnaire**
3. **Instruction**
4. **Task**
   The order of conditions is counterbalanced using a Latin square.
5. **Post-study questionnaire**

Task

Text entry as quickly and accurately as possible.

Sentence set

Sets of short sentences in Kana.

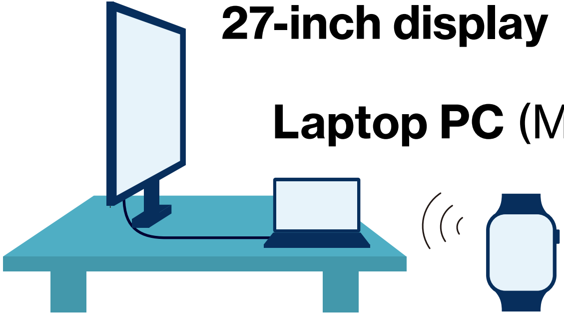
e.g. • "はるはわかれのきせつ" (Study 1)
• "じだいはまわる" (Study 2)

- **Practice**
  12 (study 1) or 10 (Study 2) sentences
- **Data collection**
  28 (study 1) or 24 (Study 2) sentences

Appendix: User Study – Metrics

Text entry speed	**Characters Per Minute (CPM)** Entered characters per minute (higher is better)
Error rate	**Total Error Rate (TER)** Number of incorrectly entered characters relative to the entire text (lower is better)
Usability	**System Usability Scale (SUS)** (higher is better)
Task load	**NASA-Raw Task Load Index (NASA-RTLX)** (lower is better)

Appendix: User Study 1 – Apparatus



27-inch display (EIZO, FlexScan EV2785)

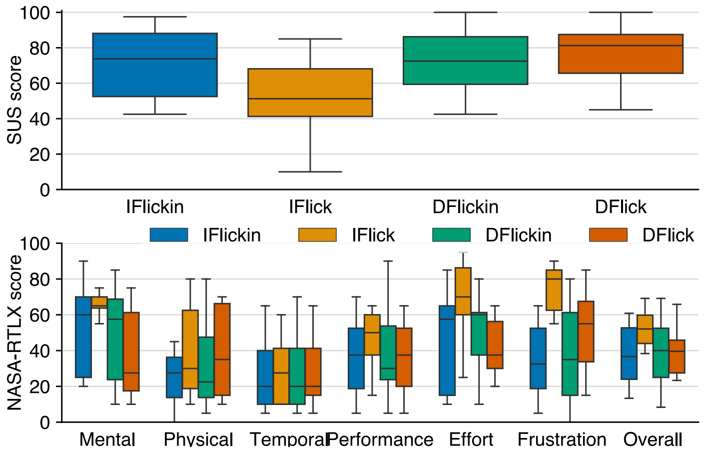
Laptop PC (MacBook Air, M1 2020)

Smartwatch
(Apple Watch Series 9)

Application

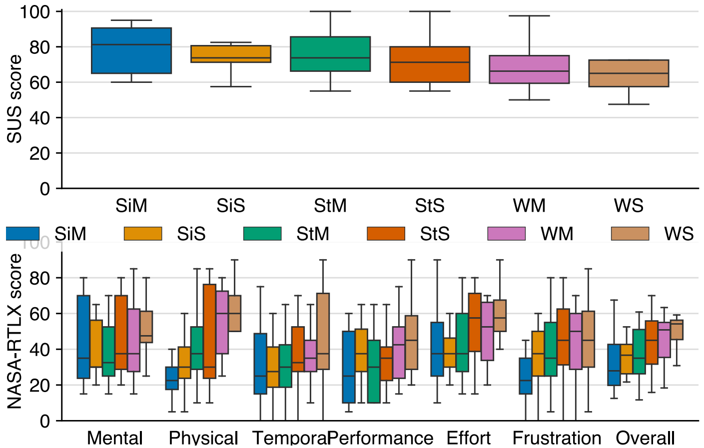
- Impementation: Swift, Swift UI
- Connection: Bluetooth Low Energy

Appendix: User Study 1 – Usability and Task load



In indirect touch,
Flick-in has
a significant higher and
a significant lower
task load than Flick.

Appendix: User Study 2 – Usability and Mental Workload



Posture

Usability

Walking has a lower than sitting.

Workload

Walking has a higher than sitting/standing.

Size

No significant difference.