Towards Remote Evaluation of Gaze Typing Systems

Hirotaka AOKI[1,2], John Paulin HANSEN[1] and Kenji ITOH[2]

[1]IT University of Copenhagen, Denmark
[2]Tokyo Institute of Technology, Japan

Remote usability evaluation

performance metrics that can be calculated automatically

Gaze Typing system

Remote Usability Evaluation of Gaze Typing

- *Remote usability evaluation* is referred to as “remotely and immediately” measuring/analysing usability-related user performance in real life condition from a distance, without any invasion, control, and limitation in use.

- Benefits of “Remote usability evaluation” in gaze typing system

Obstacle 1. Time needed for analysis

Great amount of performance data under real life condition

Obstacle 2. No control in use
No target texts

Identifying usability problems in daily use (NOT experiment)

Not invasive

Obstacle 3. Privacy
No texts obtained

Analyser/Designer

Users

Immediate detection of problems, advices / supports for a particular user, possibly

---

Performance metrics that can be calculated “quickly” (or automatically) even if obtained log-data are “incomplete” (caused by Obstacles 2 &3) are strongly required

Applicability of metrics for remote usability evaluation and new 4 metrics

	Controlled experiment	Real life condition (+Controlled experiment)
Measures that require gaze tracking	Rate of Premature Movement Errors (Aoki et al, 2005) Cumulative Deviation from the Most Efficient Scan-Path (Aoki et al, 2006)	Number of Read Text Event Per Character (Majaranta et al, 2004) 1. Attended-Keys-Not-Selected Rate
Measures that do not require gaze tracking	Error Rate (e.g., MacKenzie, 2002) Overproduction Rate (Hansen et al, 2004) Minimum String Distance (e.g., Soukoreff and MacKenzie, 2001) Cost per Correction (Gong and Tarasewich, 2006)	Words Per Minute (e.g., MacKenzie, 2002) Key Strokes Per Character (e.g., MacKenzie, 2002) Rate of Backspace Activations (Itoh et al, 2006) 2. Micro- WPM 3. Average duration of uninterrupted text-input 4. Deleted characters

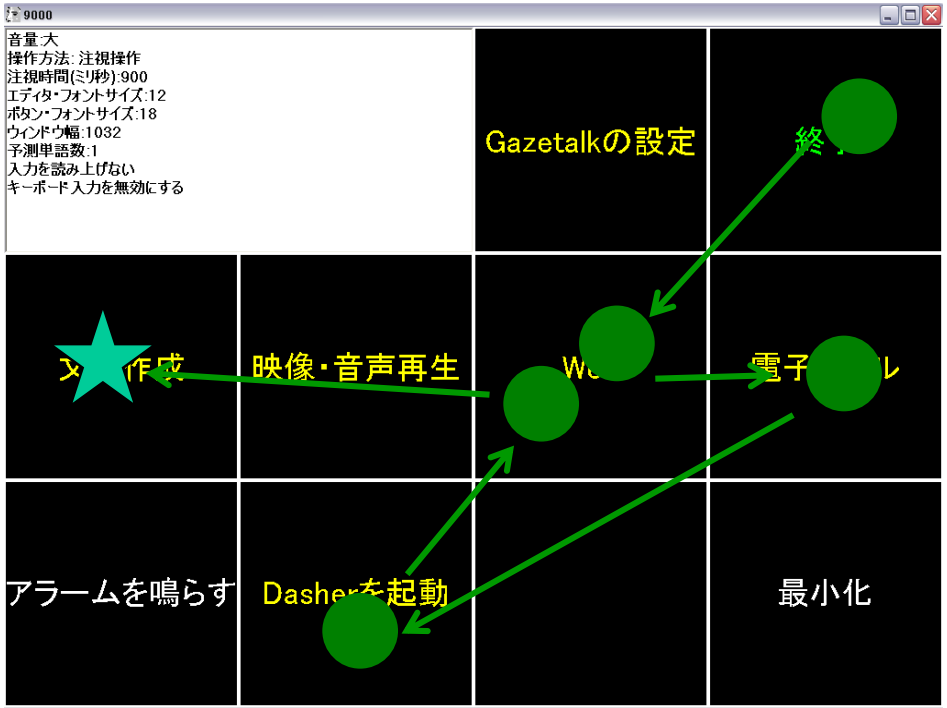
- Target texts are needed to calculate the metrics
- In general, time consuming

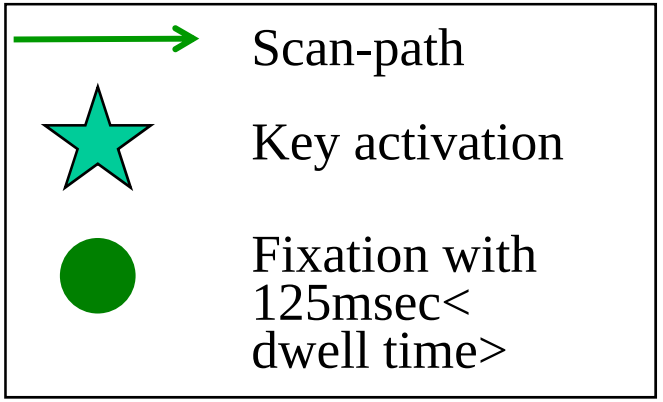
Promising candidates for remote usability evaluation

Definition of a metric ”Attended-Keys-Not-Selected rate (AKNS)”

- A metric representing a degree of how many keys the user looks at but not selects
- Calculated as number of attended-keys-not-selected divided by number of characters typed

  => Directly relating to “*1. mean number of information attended, possibly unnecessarily*”, and “*2. risks to commit wrong selections*”





Number of attended-keys but-not-selected: **5**

Comparison of AKNS with Error rate (ER)

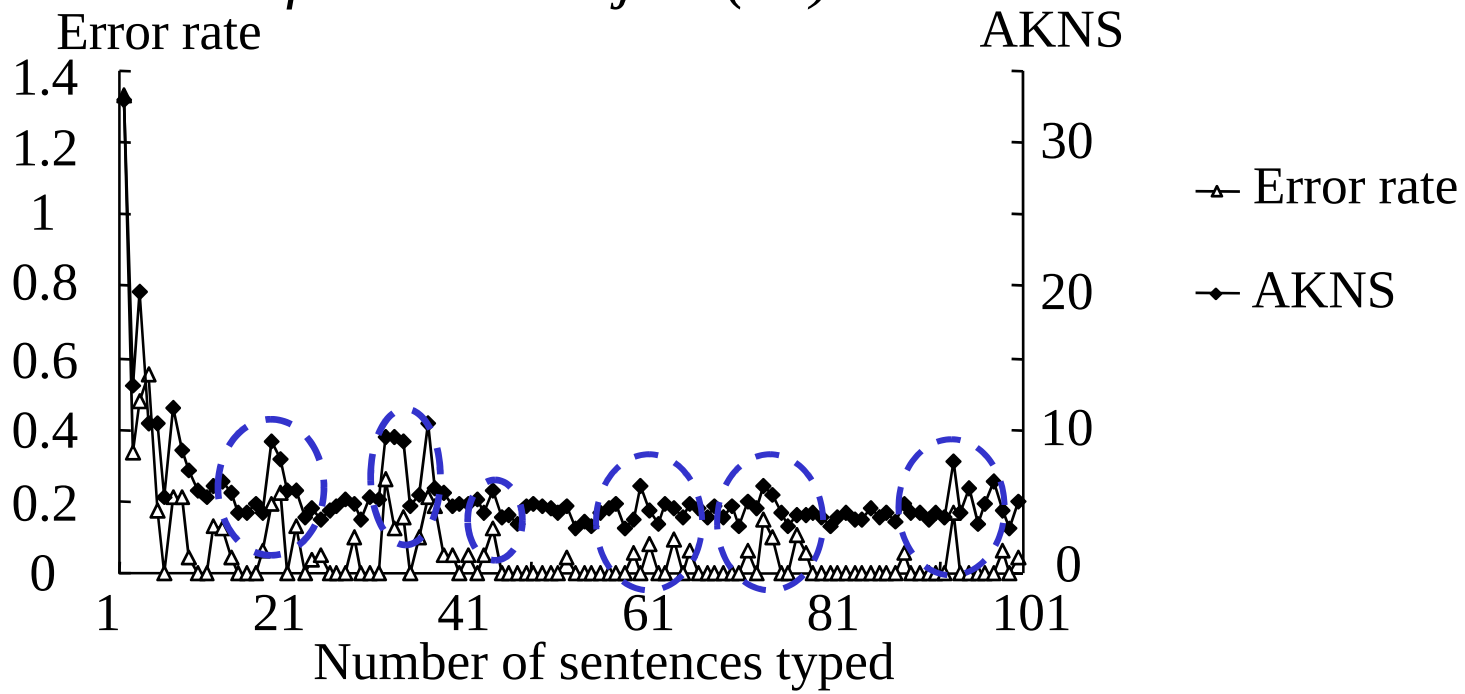
<Error rate in this study>

Number of errors divided by number of characters typed

Counted on each key activation base, and identified by close examination of all key activations

"Ground truth" of errors actually committed for each subject

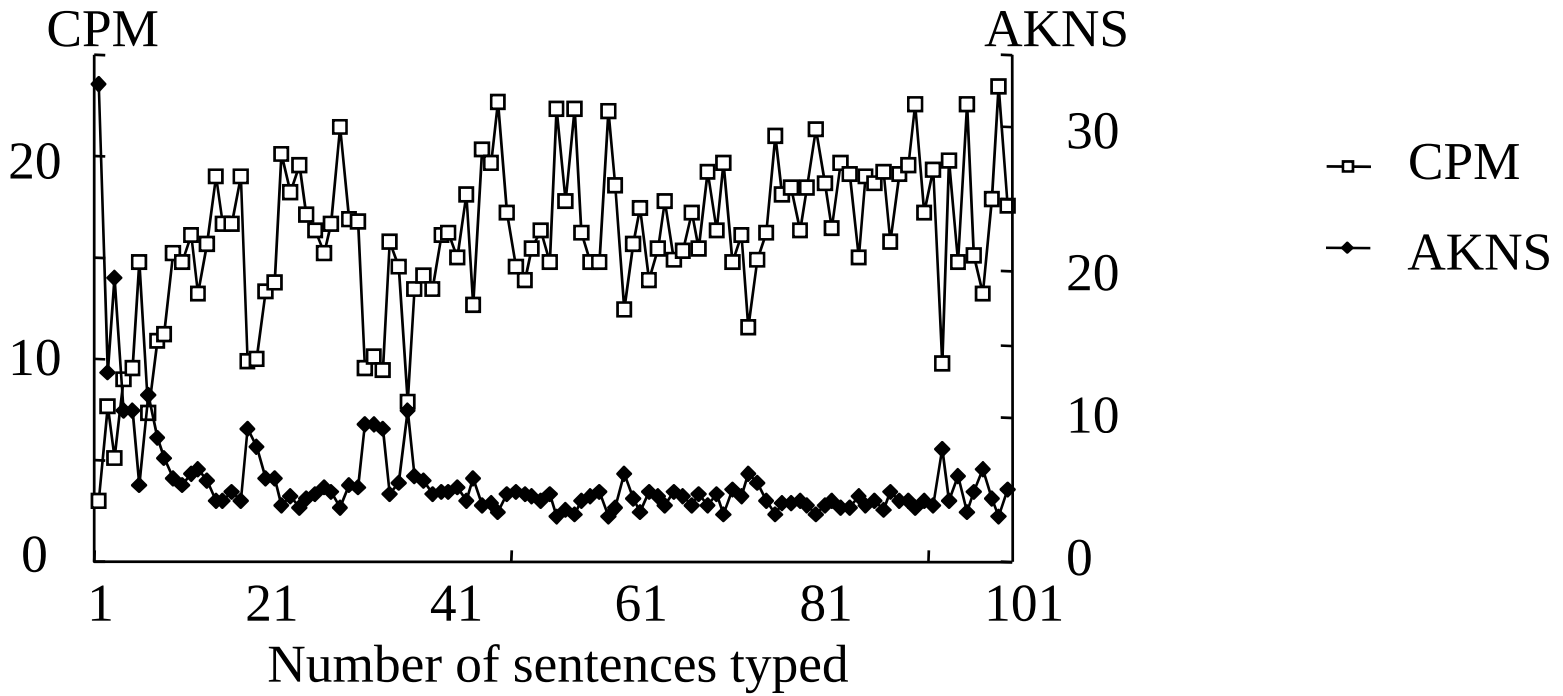
<Metrics obtained from one subject (S1)>



- Sensitive to errors

Comparison of AKNS with Characters Per Minute (CPM)

<Metrics obtained from one subject (S1)>



- Not sensitive compared to Error rate

→ Acceptable, since AKNS focus on “information attended, possibly unnecessarily”, and “risks of committing errors”

Correlations of AKNS with ER, CPM, and KSPC

<Correlation coefficients of AKNS, CPM, and KSPC with ER obtained from six subjects (S1~S6)>

	S1	S2	S3	S4	S5	S6	Mean
AKNS - Error rate	0.93**	0.45**	0.87**	0.92**	0.76**	0.91**	0.81
CPM - Error rate	-0.66**	-0.44**	-0.61**	-0.61**	-0.58**	-0.47**	-0.56
KSPC - Error rate	0.91**	0.38**	0.67**	0.88**	0.68**	0.64**	0.70

AKNS shows the higher correlations with Error rate than CPM and KSPC

- It seems that AKNS may become an attractive substitute for the error rate, since it is keenly sensitive to errors, and the most importantly, it can be calculated automatically, remotely and without knowledge of the target input stream.

Conclusion & Future Work

<*Conclusion*>

-Four metrics, that are expected to be easily calculated in remote usability evaluation condition, were proposed
″Attended-Keys-Not-Selected rate (AKNS)″, Micro-WPM, Average duration of uninterrupted text input, and Deleted characters

- ″Attended-Keys-Not-Selected rate (AKNS)″ was experimentally evaluated its validity by comparing with ER, CPM, and KSPC.

<*Future work*>

- Like AKNS, all the proposed metrics will be evaluated by use of performance data obtained in new experiments