Better User Interfaces for Occasionally Failing Technologies

Wolfgang Stuerzlinger

SCHOOL OF INTERACTIVE ARTS + TECHNOLOGY

SFU

1

We use unreliable systems as a central means to interpret input/actions

2

I still owe you monkeys.Don't let me forget!

*** Moneys. damn that autocorrect

Please pay me in monkeys! I would rather have monkeys!

WE CAN DO IT! I have faith in us!

Yeah we have to pray hardcore and rely on Godzilla on this one.

* God. Sorry Autocorrect

3

Not swords!

Not what I said

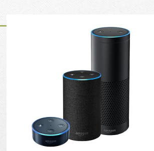
I said

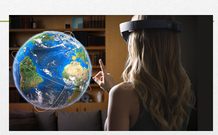
Not a worry!

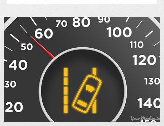
LOL - gotta love auto-correct

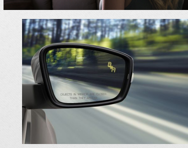
4

Other examples









5

Overreliance on Automation

Happy Birthday to you ! Happy Birthday to you! Happy Birthday dead husband! Happy Birthday to you!

Thanks. I assume you meant "dear."

Ahhhhh

Yes!!!! I mean that is a crazy autocorrect! Sorry babe.

6

Fix?

- Better technology
- Yes, **BUT**
  - Language is ambiguous
  - Gestures are ambiguous
  - World is (too) varied
- Human-in-the-loop required
- Potential legal issues

7

Complementary approach

- Better understanding of human interaction with unreliable systems
- Study perceptual, cognitive & physical aspects
- Create new UI technologies

8

Fixing errors - simple?

- Notice error
- Decide if it is worth fixing
  - Ecological rationality
- Figure out how to correct
- Correct it

Errors can happen at every step!

Errors on errors ☹

9

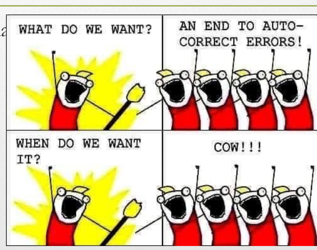
Analyze human behaviours around occasionally failing systems

- Text entry
- Gesture recognition

10

Solution?

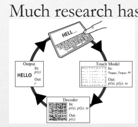
- Partia



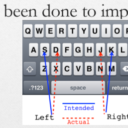
11

State-of-the-art Auto-correction

- Much research has been done to improve text entry, e.g.:



Weir et al. (2014): Touch model



Goel et al. (2012): Walking model

$$P(w|c) = (1-\lambda)P_A(w|c) + \lambda P_B(w|c)$$

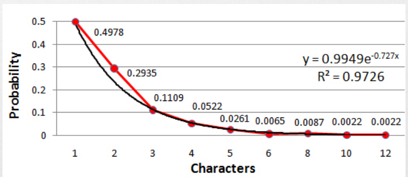
Fowler et al. (2015): Personalized LM

- However, even best approaches have error rates of ~5%
  - Higher for specific individuals and situations
  - Cannot be reduced to 0% due to ambiguity of language

12

Model of effect of text entry errors

- How should errors affect correction times?
- Probability of noticing errors (only Backspace)



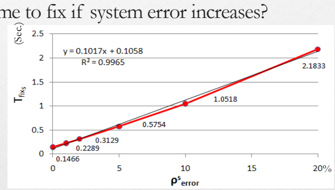
[Arif 09]

13

13

Model of effect of text entry errors

- Time to fix if system error increases?



- Nonlinear due to errors on errors

[Arif CHI 10]

14

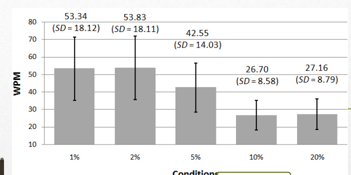
14

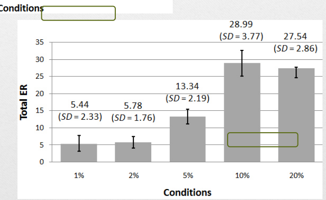
Experiment

- Text entry with "faulty" keyboard
  - Adjacent key with controlled failure rate
  - 1, 2, 5, 10, 20% errors

15

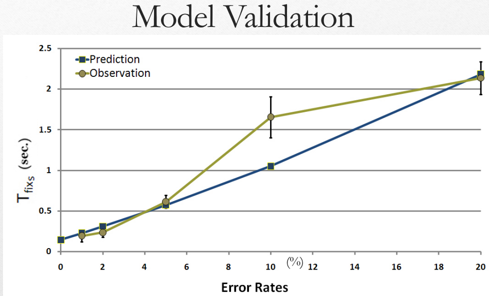
15





16

Model Validation



18

18

Adaptation

- Errors can happen repeatedly
- Humans can adapt

- Rely on human adaptation?
- OK if technology is predictable

- BUT…

19

Adaptation: Core problem

- Technology not always predictable
  - Recognition tech sensitive to ”random” variations
  - Changes due to updates/upgrades/…
- People don’t generally understand underlying systems
- Underlying system appear random
  - So we cannot predict if & when they will fail
  - Can’t adapt to failures

20

Research Questions

- Do users adapt to a faulty system?
- What influences this adaptation process?

[Arif GI 14]

21

21

Research Questions

- Do users adapt to a faulty unistroke gesture recognizer?
- What influences this adaptation process?

22

22

Research Questions

- Do users adapt to injected misrecognition errors of a unistroke gesture recognizer?
- What influences this adaptation process?

23

23

Research Questions

- Do users adapt to injected misrecognition errors of a unistroke gesture recognizer by switching to an alternative gesture set?
- What influences this adaptation process?

24

24

Errors in Gesture Recognition

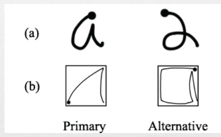
- Generally error-prone
  - Useful when at least 97% accurate [LaLomia, 1994]
  - Abandoned when below 40% [Karam & schraefel, 2006]
- Usually compare performed with existing
  - Misrecognition error
    - Most common
  - Failure to recognize (! library, accidental strokes)

25

25

Alternative Gestures

- Many support several drawing variations



- Alternatives:
  - Less intuitive
  - Harder to discover

26

26

Apparatus

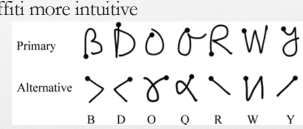
- $1 unistroke recognizer [Wobbrock et al., 2007]
  - 7 templates / letter
  - 99% accuracy rate (.7% misrecognition, .3% failure to recognize)
- Multistroke allows many variations
  - Difficult to identify human errors
  - Users require time to identify an issue

27

27

Supported Gestures

- Graffiti more intuitive

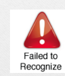


- Graffiti ≈ Unistrokes [Castellucci,MacKenzie, 2008]
  - Method switch will not compromise performance

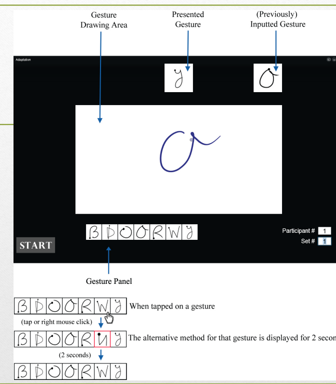
28

28

- Discoverability
- Error handling

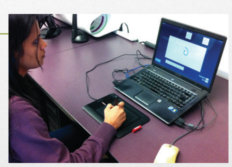


- Synthetic misrecognition



29

Study 1 – 0-30%



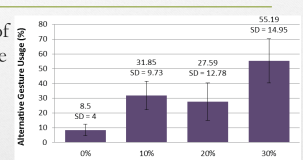
- 12 novice participants
- Practice: 5/Graffiti gesture
- 7 letters – 630 times
  - 3 random Graffiti with 10, 20, 30% injected misrecognitions
- Alternative gesture use was not forced
  - No error injection on 1+ attempts

30

30

Alternative Method Usage

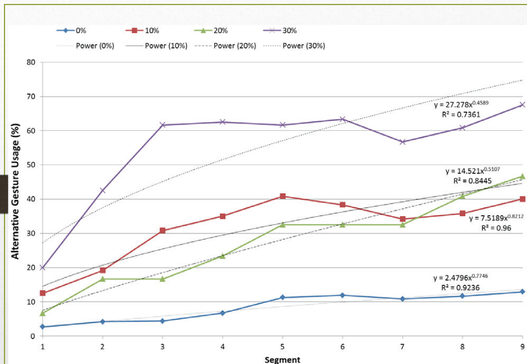
- Significant effect of misrecognition rate



- 0, 10-20, 30% significantly different

31

31

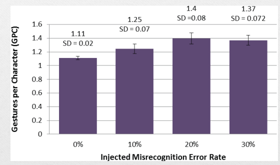


32

32

Time & GPC

- Input time – no effect
- Gestures per Character –effect

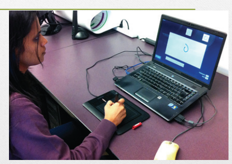


33

33

Study 2 – 0-10%

- 12 novice participants



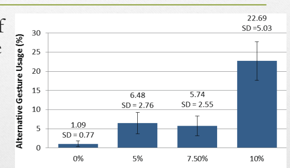
- Same as Study 1
- 7 letters – 630 time,
- 3 random Graffiti
  - with 5, 7.5, 10% injected misrecognitions

34

34

Alternative Method Usage

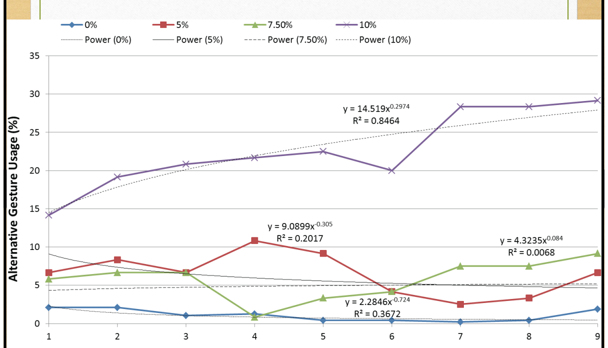
- Significant effect of misrecognition rate



- 0, 5-7.5% and 10% significantly different

35

35

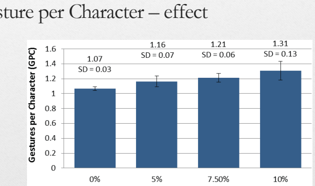


36

36

Time & GPC

- Input time – no effect
- Gesture per Character – effect



37

37

Outcomes & Recommendations

- Users can adapt to a faulty gesture recognizer
- Adaptation depends on injected error rate
  - Similar trends in psychology, skill acquisition, UI
    - Greater effort = more recall-based actions
- More than 90% accuracy rate is necessary
- Users must have options
- Alternates should be easy to discover

38

38

Further Discussion

- Some adaptation for 0% as well
- Half did not identify all 3 faulty letters ☹
  - Or did not spend effort to learn
  - Different cognitive strategies / personalities?

39

39

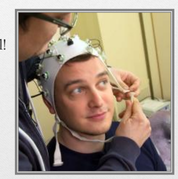
Adaptation

- Do humans adapt?
  - Yes
  - BUT
    - Only to things they *notice*
    - Sufficiently frequently
    - And reliably
  - ALSO
    - Benefit needs to be high enough

40

Is noticing errors enough?

- Sense reaction to errors
  - Error-Related Potential in EEG signal!
- Brain-computer interface to sense user reactions to incorrect auto-corrects
- (Trigger better system responses by offering different corrections)



41

Self-Repairing Auto-Correction

„I would live…"

I would libe | Send → auto-correction → I would like | Send

Error detected
→ trigger repair attempt

I would live | Send
like
lice
line

→ Reduced effort for repairs

- Step 1: Detect reaction to error

42

Error Potentials

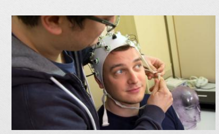
- How can we hope to detect auto-correction errors?

„I would live…"

Perceived mismatch

I would libe | Send → I would like | Send

- Perceived mismatch results in characteristic brain activity (error potentials)
- Capture through EEG
- Goal: automatically detect error potentials triggered by auto-corrections

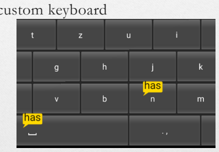


[Putze ICMI '17]

43

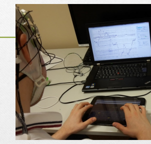
Experimental Paradigm

- Android tablet text entry app with custom keyboard
- Dictionary-based auto-correction
  - Select replacement randomly from entries with minimal edit distance
- User has to notice errors → Draw attention to auto-corrections
  - Audio-tactile cue
  - Multiple visual cues at potential gaze targets
- Rigged keyboard (forced 5% switched characters)
  - Increase number of correction events in limited time frame

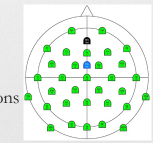


44

Data Collection

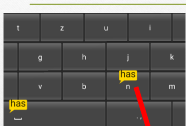


- 12 participants
- Typed 120 sentences (+15 training) for about 23 minutes
  - Sentences from OpenSubtitles phrase set
- Questionnaire on typing behavior and self-assessment
  - Validated people noticing auto-corrections
- Recorded EEG data at 32 electrode positions + synchronized user's typing behavior



45

Classification Approach



For each auto-correction, classify data based on correction success

EEG data → error (26%)

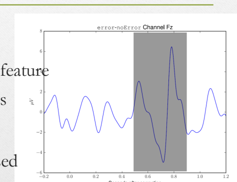
typing data → noError (74%)

1s

46

EEG Features

- For each EEG electrode:
- Use filtered, down-sampled EEG signal as time-domain feature
- Use power spectral density as frequency-domain feature
- Both types of features carry information about event-based EEG patterns
  - May be contaminated with artifacts: user moving, gazing, typing



47

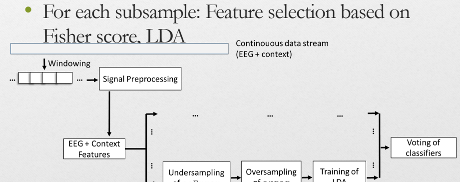
Context Features

- User behavior and input characteristics contain relevant information
- Two types of context features:
- 1) Encode likelihood of auto-correction error
  - Length of the replaced word (# characters)
  - # candidate words of minimal edit distance
- 2) Encode likelihood of user perceiving error
  - Typing speed for replaced word relative to average typing speed
  - Time before user continues typing during EEG window (in ms)

48

Classification

- Data set unbalanced (#noError > #error)
  - Handle by oversampling & undersampling + bagging
- For each subsample: Feature selection based on Fisher score, LDA



49

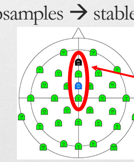
Classification Results

		Accuracy	Precision	Recall	F1 Score
Baseline	Mean	0.74	-	-	0.38
EEG	Mean	0.69	0.25	0.38	0.30
	SD	0.08	0.10	0.13	0.07
Context	Mean	0.81	0.40	0.70	0.49
	SD	0.03	0.19	0.25	0.18
Combined	Mean	**0.85**	**0.82**	**0.65**	**0.72**
	SD	0.03	0.05	0.11	0.08

50

Feature Interpretation

- For every person, ~6 features were selected in 75% of all folds and subsamples → stable intra-personal
- Across all persons, ~16 features selected in more than 40% of all folds and subsamples → stable inter-personal
- Plausible features:
  - Does not rely on ocular artifacts



51

Discussion

- Auto-correction errors can be detected!
  - From EEG and context features
- Step 2: Work in progress
  - Challenge: classification window alignment
    - Delay in perception?
    - Include eye tracking?
  - Good enough to improve text entry efficiency?

52

Better UIs for Occasionally Failing Technology

- Better understanding of human interaction with unreliable systems
- Study perceptual, cognitive, & physical aspects
- Create new UI technologies

53

Bigger Issues

- Can we eliminate all failures?
  - World extremely complex
  - Legal issues
- Overreliance on automation
  - Automation bias
- Ecological Rationality
- Misperception of probabilities

54

Future work

- Improvements to text entry [Alharbi GI 19]
  - Better error visualizations & correction methods
- EEG-based active auto-correct
- Auto-correct in scheduling
- Voice recognition correction
- Address misperception of probabilities
- …

55

Thanks

- Students
  - Ahmed Arif
  - Euclides Mendonca Filho
  - Benedikt Iltisberger
  - Andriy Pavlovych
  - Rebecca Singh
  - Ohoud Alharbi
  - Aunnoy Mutasim
- Collaborators

NSERC CRSNG

Graphisme, animation et nouveaux médias
GRAND
Graphics, Animation and New Media

56

57