Experiment

Participant: Patrick Welche
User: 3

All results are correct except where indicated.

Ticker

Session 1:
Duration: 1 hour: 1x Tutorial: 40 minutes, 1x Test session: 15 minutes:

Sub-session 1: Speed: 0.65, 10 minutes, Tue 4 March.
yes_.
done_.

- Wrote “jones_” and time-out error on “.”.

for_.
large_.

Sub-session 2: 30 minutes, **Speed:** 0.7, Tue 4 March.
the_quick_brown_fox_jumps_over_the_lazy_dog_.

- quick_ (tried undo many times, and then got out of sync with one letter), and fox (wrote for even though x was well highlighted).

Sub-session 3 (Testing): 15 minutes, **Speed:** 0.7, Tue 4 March.
she_is_my_best_friend_.
why_do_you_say_that_so_.

- All correct but got confused on do_ and requested a pause to ask what the options are, then started over with word by waiting for new word.

this_should_not_to_happen_to_anyone_in_the_world_.

Sub-session 4 (Tutorial): 30 minutes, **Speed:** 0.7, Thu 7 March.

- Went through rules again.

done_.
the_quick_brown_fox_jumps_over_the_lazy_dog_.

- Brown_ , fox_ (for_), jumps_ (time out, started new word and was then out of sync).

Sub-session 5: 15 minutes, **Speed:** 0.7, Thu 7 March.
no_more_headaches_.

- no_ (so), more (time out, letters were out of sync, after going through it more than once).

i_will_read_it_to_you_.

- it_ (in_)

i_would_like_some_soup_for_lunch_.
saw_your_keys_by_the_sink_.

Sub-session 6: 15 minutes, **Speed:** 0.7, Thu 7 March.

- This was accidentally recorded as part of sub-session 5 and copied into sub-session 6 afterwards. The log file can therefore be found in sub-session 5.

what_time_is_it_.
take_care_.
i_enjoy_the_music_.
turn_on_the_tv_.
i_cant_wait_for_the_pain_to_go_away_though_.

- cant_ (time out, undo on “a”)

Session 2: 90 minutes, **Speed:** 0.7-0.79, Fri 8 March.
Sub-session 1 (Tutorial): 15 minutes, **Speed:** 0.7, Fri 8 March.
yes_
done_.
fox_ (practice mode & calibration mode)

- Look at output to see click for “x”.

the_quick_brown_fox_jumps_over_the_lazy_dog_.

- quick_ (quite_), brown_(triathlon_ - went through once, got confused and went immediately on to fox_), fox_(for_). Two spaces after jumps_, but I did edit a file (see bug.txt) while the user clicked, and it seemed as if some clicks then didn't go through while writing over_ (only time when I tried to do something else while the user clicked).

Sub-session 2: (Testing): 15 minutes, **Speed:** 0.7-0.73, Fri 8 March.
you_got_to_figure_it_out_. (0.7).
could_you_do_it_for_me_. (0.7).
you_are_great_. (0.72)
where_are_you_going_. (0.72)
i_love_you_. (0.73).

- Undo “l” on “love”.

wishing_you_all_the_best_on_your_birthday_. (0.73).

- Stopped session at “all_”.

Sub-session 3: (Testing): 15 minutes, **Speed:** 0.7-0.76, Fri 8 March.
water_. (0.73)
i_wish_i_could_. (0.73)
you_are_a_true_friend_. (0.74)
you_go_and_rest_. (0.75).

- “go_g”.

find_the_tv_remote_for_me_. (0.75)
why_not_. (0.76)
feeling_better_now_.   (0.76)

i_need_to_buy_cat_food_. (0.76)

- need_ (needed_)

Sub-session 4: (Testing): 15 minutes, **Speed:** 0.77-0.79, Fri 8 March.
can_you_walk_the_dog_for_me_. (0.77).
please_help_me_miss_. (0.77).

- Had to repeat help twice.

what_is_the_time_. (0.78)
could_you_please_massage_my_shoulder_. (0.78)

- shoulder_ (should_)

how_have_you_been_. (0.79)

- you_ (time-out error)

when_are_you_coming_back_. (0.79)

Session 3: 90 minutes, **Speed:** Mainly 0.77 started with 0.8, Fri 8 March.

- The session started a bit wonky at speed 0.8 and 0.79 in sub-session 1(about one error in every phrase). The speed was then set to 0.77 (re-calibrated the click distribution) and was kept at that speed. It went a lot better, but not as good as session 2. We did go on directly from sub-session 4 in previous session (with a short break), so the user might have gotten a bit tired.

Sub-session 1 (Tutorial): 10 minutes, **Speed:** 0.79, Fri 8 March.
the_quick_brown_fox_jumps_over_the_lazy_dog_.

- fox_ (for_), lazy_ (time-out error) (0.79)

Sub-session 2: (Testing): 15 minutes, **Speed:** 0.7-0.73, Fri 8 March.
it_hurts_a_bit_on_my_right_side_.

- Wrote it_hurts_ (time out on it_ and hurts_, speed 0.8). The user really seemed to struggle, so the speed was set to 0.79, and the click distribution was re-calibrated with “yes_” at speed 0.79.
- Then wrote whole sentence (0.79) with time-out error on right_.

i_would_like_a_cake_for_my_birthday_.

- birthday_ (another_) (0.79)

i_prefer_the_other_one_.

- other_(the), one_(other_): The user got confused, even with the words represented to him in a voice prompt. (0.79)

i_like_to_go_places_.

- places_(that_) (0.79)

Sub-session 3: (Testing): 15 minutes, **Speed:** 0.77, Fri 8 March.

- Because the user seemed a bit exhausted, it was decided to reduce the speed slightly (with calibration on yes_). The speed was then constant for sub-session 3 + 4 (no re-calibration).

do_you_think_mom_will_come_.

- mom_(time-out error)

bath_time_.
why_.
for_you_that_is_great_.
i_bought_the_groceries_yesterday_.
are_we_going_out_today_.
which_day_.
shall_i_take_you_out_for_a_walk_.

- (Session stopped after “take_” was written).

Sub-session 4: (Testing): 15 minutes, **Speed:** 0.77, Fri 8 March.
can_i_borrow_some_money_.
what_is_on_tv_.

- is_(time-out error)

i_need_to_refresh_.
i_wish_i_could_join_you_.
take_care_.

- care_(time-out error)

somebody_answer_the_door_.
wish_you_all_the_best_.
why_.
pee_.

- (time-out error).

Session 4 80 minutes, **Speed:** 0.7 - 0.86, Fri 8 March.

- Before each speed setting the user calibrated with yes (should also be done in Grid, i.e., user should “practice” with “yes_”).

Sub-session 1 (Tutorial): 10 minutes, **Speed:** 0.7, Fri 8 March.

- All_. (correct)
- done_. (correct)

Sub-session 2: (Testing): 15 minutes, **Speed:** 0.7-0.74, Fri 8 March.
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.7)

- fox (time-out error)

the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.72)
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.74)

- fox (for_)

Sub-session 3: (Testing): 15 minutes, **Speed:** 0.76-0.8, Fri 8 March.
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.76)

- the_(time-out, beginning of sentence), fox (time-out error)

the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.78)

- fox_(time-out error)

the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.8)

- lazy_(hazy_)

Sub-session 4: (Testing): 15 minutes, **Speed:** 0.82-0.86, Fri 8 March.

the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.82)

- fox (for_)

the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.84)

- brown_(time-out, thought it was a “g” instead of a “b”), fox_, jumps_, dog_ (all time-out errors)

the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.86)

- fox_(for_), quick_, brown_, jumps_, lazy_, dog_ (time-out errors)
- user noted that 0.82 was very comfortable, but breaking point appeared suddenly, and that click times might be learned.
-

Session 5 Noisy experiment, Tue 26 March.

Sub-session 1 (Tutorial + testing): 30 minutes, **Speed:** 0.65, Fri 8 March.

- 10 minute tutorial (in tutorial mode)
- Only testing (15 minutes) results reported, noise settings
- Gauss delay = 0.8, sigma =0.05, fp_rate, fr=0
- Calibrated (with learning rate 1.0). Then fixed fp_rate to 0.25 and fr to 0.1 (no learning).
- End delay = 1.55 seconds

pass_me_that_book_there_.
wash_the_dishes_.

- was_

dying_of_homework_exhaustion_.
maybe_next_week_.
my_skin_is_red_and_stinging_.

Sub-session 2 (Tutorial): 15 minutes, **Speed:** 0.65, Fri 8 March.

- Gauss delay =1.5, sigma=0.05, fpr_rate=1/3, fr=0.1
- 5 False positives were generated (which seems right), as counted manually from the log file.
- About 25 false rejections were generated out of 193 true clicks (198 clicks in total of which 5 were false positives), which seems in the right ball park (13%).

good_job_.
we_have_been_doing_a_lot_of_overtime_.
did_you_bring_some_vegetables_.
when_did_you_call_me_.
what_is_the_problem_.

General notes:

- The user found it very hard in the first session to keep track of the words in the phrases and do the experiment, forcing him to look at the “phrase cribsheet” too often. It was then decided to include a voice prompt telling the user which word to write next, whenever a new word had to be written. By the middle of session 2 the user noted that he hardly looked at the crib sheet.

- The user found the letters “a” and “l” difficult to write – he managed but it was always difficult. “l” was easier than “a” as the “l” at the end of the channel is very easy to select – this was also observed by the supervisor (looking at the letter likelihoods all the time). The supervisor noted that “r” also seems difficult to select sometimes.
- The user noted that it might make it easier to direct his attention from the beginning of the sequence of the silence at the end was proportional to the period of the channel rhythm, making it easier not to loose the rhythm in the first place. Eventually the silence at the end should actually be completely removed – it wastes time. However, training should be fast enough and the code will have to be adapted significantly.
- The user noted that the system is more lenient on longer words than shorter words. Undo is therefore typically only used when writing a short word (2-4 letter
- words).

Grid

Session 1 70 minutes, **Speed:** 1 second to 0.5 seconds (scan delay), Sat 16 March.

- User reached 0.6s scan delay easily, made some mistakes with 0.5s. There seems to be a clear breaking point at 0.4s (only tried very quickly at last 3 minutes of experiment). During the next session the user could easily select letters at 0.4s.
- User said that he always clicked between ticks (actually this is correct, doesn't need to change tick sounds at all). Fastest mode is when following the rhythm of the ticks, i.e., anticipating.
- User did double click on some of the vowels in the beginning – a real user will find this difficult to do with a real switch, especially because the grid usually has a "minimum time" between clicks. Asked the user explicitly not to do this from session 1, sub-session 3 onwards.
- All sentences correct, except where indicated (at scan delay of 0.5s).

Sub-session 1 (Tutorial): 10 minutes, **Speed:** 1-0.9 (seconds), Sat 16 March.

- Yes_. (e – typed a – deleted a – typed e) (1s)
- all_. (1s)
- done_. (1s)
- the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.9s)
  - First the_: min number of scans used.
  - quick_(u → wrote v → deleted → wrote u
  - jumps(m → skipped 1 iterations, otherwise minimum number of letters).

Sub-session 2: (Testing): 15 minutes, **Speed:** 0.8-0.6 seconds, Sat 16 March.

all_the_best_. (0.8s)

- all_(min number of scans)

we_all_will_go_together_. (0.8s)

- will_ (_ → wrote e → delete → then write _)

how_much_. (0.7s)

was_tasty_thanks_.(0.7s)

i_need_my_sweater_. (0.6s)

Sub-session 3: (Testing): 15 minutes, **Speed:** 0.6-0.5 seconds, Sat 16 March.

do_you_know_the_time_of_her_flight_. (0.6s)

- paused at "k" of known

are_you_safe_at_home_. (0.5s)

- Wrote arc_ instead of are_
- at_(atf → delete f → wrote_)

this_device_is_my_voice_. (0.5s)

- this_(s → undo row), wrote thjs_
- device_(c → undo row)

Sub-session 4: (Testing): 15 minutes, **Speed:** 0.5 seconds, Sat 16 March.

give_me_my_medicine_dear_. (0.5s)

- my_(myd_)
- medicine_(first i → wrote j → delete → i)

thanks_for_dropping_by_. (0.5s)

- thanks_(k → undo)

what_time_. (0.5s)
are_you_having_trouble_reading_my_message_. (0.5s)

- Stopped at my_.

Session 2

Sub-session 1 (Tutorial): 15 minutes, **Speed:** 0.4 seconds
the_quick_brown_fox_jumps_over_the_lazy_dog_.
are_
are_you_having_trouble_reading_my_message_.

Sub-session 2: (Testing): 15 minutes, **Speed:** 0.4 seconds
we_are_eating_burgers_.
i_love_math_.
how_did_you_know_her_.

Sub-session 3: (Testing): 15 minutes, **Speed:** 0.3 seconds
do_you_know_when_the_mail_comes_.

- time-out error on "."

really_enjoyed_your_company_.

- enjoyed: lots of undo and delete

talk_to_you_again_tomorrow_.

- talk_(talkg_)

which_button_do_i_press_.

Sub-session 4: (Testing): 15 minutes, **Speed:** 0.4 seconds
i_really_love_you_guys_.
will_you_come_with_me_.
you_are_amazing_.
nice_place_.
just_saying_hi_.

Session 3

- Speed-saturation experiment. (The quick brown fox ...)

Sub-session 1 (Testing): 20-25 minutes, did take short breaks in between.
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.5s)
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.4s)
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.3s)
the_quick_brown_fox_jumps_over_the_lazy_dog_. (0.2s)

- the_(rhj_), quick_(rujek_), brown(bxown_)
- breaking point.

Session 4 40 minutes, **Speed:** Very slow: Noisy experiment, Friday 22 March

- During this session we tested two noise settings, by synthetically adding Gaussian noise to the user's click times and by generating a small number of false positives: 1) a very modest setting (no false positives / negatives), only Gaussian noise (mean=0.8s, sigma=0.05s), and two a more unreliable switch, which can easily be found in practice (fp_rate=0.3 per minute, fr=0.1, mean=1.5s, sigma=0.05).
- In all cases the scan delay was set (delay=mean + 3*sigma + 0.4s). The user wrote comfortably at 0.4s and the 3*sigma is to make sure the clicks at the boundaries are safe.

Sub-session 1: (Testing): 15 minutes, **Speed:** 1.4 seconds, Friday 22 March.
can_you_get_me_a_blanket_please_.

- Ignored “t” of blanket (user did click twice on it and it was ignored)

good_luck_.
Sub-session 2: (Testing): 15 minutes, **Speed:** 2.1 seconds, Friday 22 March.
Noisy experiment: fp_rate = 1/3, fr=0.1, gauss_mean=1.5s, sigma=0.05

you_know_this_is_not_the_first_

- know_(knowba_)
- User did notice false negatives, and that false positives were really frustrating.
- False positives are generated when starting a new scan. Going through log file, there were 6 false positives in total: 2 were generated while reading the voice prompts (and therefore ignored), 3 were generated while writing know_, and one was generated after the true click (i.e., user didn't get to it).
- About 9 false rejections out of 70 which is about 13% (seems to be the right ball park).