

Deliberate or spontaneous? A duration reject window for webcam blink typing, checked on public blink videos and one recorded session

Saksham Sharma

BITS Pilani

India

2025eb01657@online.bits-pilani.ac.in

Abstract

Blink-based text entry uses the eyelid as a key, but the eyelid also closes on its own many times a minute. MorsEye is an open-source browser system that types Morse code from eye closures on a laptop webcam: it measures closures from MediaPipe face landmarks and blendshape scores, ignores closures shorter than a reject window, and maps longer ones to dots and dashes. A preregistered single-person study was planned, but data collection stopped after one session without labelled closure times, so its confirmatory tests were not run. We report what remains. On 408 annotated natural blinks of 4 people in a public video dataset (Eyeblink8; secondary, not blind), a 200 ms reject window cut false triggers from 10.5 [7.5, 13.2] to 4.1 [2.7, 5.7] per minute with a blendshape detector, but only from 11.8 [9.2, 14.2] to 10.2 [7.5, 12.9] with the eye aspect ratio. In the one recorded session (exploratory), the author’s cued short closures mostly lasted longer than the 600 ms dot limit (a boundary near 960 ms would have classified 91% of short and long responses correctly, against 56%), and winks closed both eyes as far as the camera could tell.

1 Introduction

People who cannot speak or use their hands may still control their eyelids, which makes blinks a candidate input for text entry [4, 5]. The difficulty is that the “key” also moves on its own: spontaneous blinks happen many times a minute, and a system that types a symbol for each of them is unusable. Webcam blink typing therefore has to tell deliberate closures from spontaneous ones, using only what an ordinary laptop camera at about 30 frames per second can see.

MorsEye takes a simple approach: deliberate closures are made longer than natural blinks, and closures shorter than a *reject window* are ignored. Short and long deliberate closures become Morse dots and dashes. This paper reports three things:

- an open-source browser system that runs entirely on the device, with per-user calibration, a duration segmenter and a study and labelling mode (Section 3);
- a check of the reject window on other people’s natural blinks in a public video dataset (Section 5);
- an exploratory account of one recorded session of the author, which shows where the default timing bands did not fit (Section 6).

The paper also reports, with the dated deviations, why the preregistered study it was designed around was not completed (Section 4).

2 Related Work

Soukupová and Čech [6] proposed the eye aspect ratio (EAR), computed from six landmarks around each eye, as a per-frame measure of eye opening for real-time blink detection. Grauman et al. [4] used blinks and eyebrow raises from video as an interface for people with severe motor impairments, and Majaranta and Riih  [5] reviewed design issues in eye typing. Annotated blink videos, such as Eyeblink8 [1] and the larger Researcher’s Night set [2], were built to evaluate blink detectors; they contain natural blinks only, not deliberate ones, which is how they are used here.

3 System

MorsEye runs in the browser (Fig. 1). Each camera frame is timestamped and passed to MediaPipe’s face landmarker [3], which returns face landmarks and blendshape scores on the device. Two signals are computed per frame: the mean EAR of both eyes [6], and the larger of the two eyeBlink blendshape scores.

Detectors. A detector marks the eyes closed when its signal crosses a threshold, with hysteresis so that noise near the threshold does not split one closure into several. Closure durations come from frame timestamps, not frame counts. Five detectors are compared: D1, EAR with the fixed threshold 0.21; D2, EAR with a calibrated threshold; D3, blendshape with the fixed threshold 0.5; D4, blendshape with a calibrated threshold; and D5, a wink detector that accepts a closure of one eye only while the other eye’s score stays below 0.3.

Calibration. The user makes ten cued short and ten cued long closures and then keeps the eyes open for 15 s. The threshold is placed halfway between the median open value and the median of the closures’ peak values, with hysteresis of 10% of that range.

Segmenter and decoder. Each closure is classified by duration. With the default bands, closures under 200 ms are ignored (the reject window), under 600 ms are dots, under 1400 ms are dashes, under 2000 ms are a “hold” (two holds in a row pause input), and longer ones delete the last character. A pause after a symbol ends the letter, and a longer pause adds a space.

Study and labelling modes. A study mode runs the protocol with full-screen cues and logs every frame’s features (never the image) with cue and block times. A local video can be recorded for labelling, but it stays on the laptop and is never published. A labelling tool shows the video next to the signal traces for marking closure onsets, offsets and types by hand.

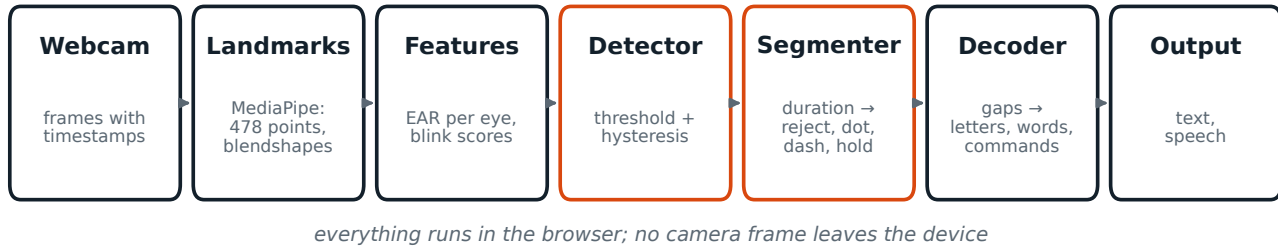


Figure 1: The MorsEye pipeline. Everything runs in the browser; no camera frame leaves the device.

4 The Planned Study and What Happened

The evaluation was preregistered on 24 Oct 2026, before the first session (docs/preregistration.md). Because no volunteers were available, it was designed as a case study of one person, the author:

- 6–8 detection sessions, each with a calibration block, a reading block with no cues, a block of 40 cued short, long, left-wink and right-wink closures, and a reading block with occasional cues;
- five days of typing practice;
- a check on public blink videos.

Confirmatory hypotheses compared calibrated with fixed thresholds (H1), blendshape with EAR (H2), and detection with and without the reject window (H3).

Data collection stopped after the first detection session, at the author’s decision. In that session the author checked the response type to each cue in the labelling tool, but did not mark closure onsets and offsets. Consequently:

- no session has labelled closure times, so H1, H2 and H3 on the author’s data were *not tested*, and the minimum of six sessions was not met;
- the typing study was not run.

These changes, and smaller ones (no pilot session; a corrected count of missing frames), are recorded with dates in the preregistration’s Deviations section. What remains is the public-video analysis, which the preregistration lists as secondary and not blind, and the single session, which is reported as exploratory.

5 Natural Blinks in Public Videos

Data and method. Eyeblink8 [1] contains 8 videos of 4 people (one wearing glasses) at home, 40 minutes in total at about 31 fps, with 408 annotated natural blinks. The app’s features were extracted from every frame with the same MediaPipe version and model as the browser, and the two fixed-threshold detectors (D1, D3) and the segmenter were run on them. Calibrated detectors could not be run, because the videos contain no deliberate closures to calibrate from. Every blink in the videos is spontaneous, so every detected closure that the segmenter does not reject is a false trigger. False triggers per minute are computed per video, with and without the reject window, and summarised as the mean over videos with a 95% percentile bootstrap CI (10,000 resamples).

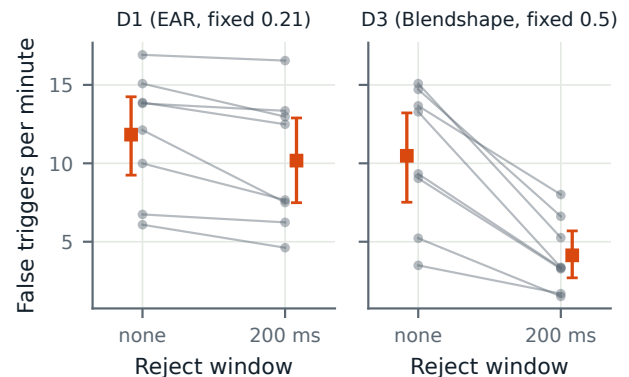


Figure 2: Eyeblink8: false triggers per minute on natural blinks, without and with the reject window. Grey lines are videos; squares are means with 95% bootstrap CIs over videos.

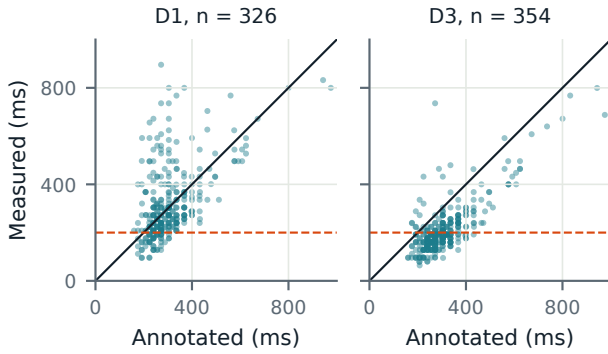
This analysis is secondary and *not blind*: before the preregistration was written, the annotated blink durations had already been computed and both detectors run on one of the videos. The blind version of this check, on the Researcher’s Night dataset [2], was not done: the author decided not to request that dataset.

Results. The annotated blinks lasted a median of 272 ms (5th–95th percentile 192–560 ms); 9% were shorter than 200 ms. Table 1 and Fig. 2 give false triggers per minute. With the blendshape detector (D3), the reject window lowered them from 10.5 [7.5, 13.2] to 4.1 [2.7, 5.7] per minute (ratio of means 0.39). With the EAR detector (D1), they fell only from 11.8 [9.2, 14.2] to 10.2 [7.5, 12.9] (ratio 0.86). For both detectors the rate fell in all eight videos (Wilcoxon signed-rank $p = 0.008$ for D1 and 0.008 for D3, the smallest possible with eight videos), but only D3 met the preregistered criterion of at least halving the rate. The four people are not a sample of a population, and the eight videos are not independent, so these intervals describe these recordings only.

Why the detectors differ. The difference comes from how long each detector measures a blink to be (Fig. 3). For blinks that both the annotation and the detector found, D3’s closed period was a median 96 ms shorter than the annotated blink (interquartile range -128 to -64 ms), because its threshold of 0.5 is crossed only near the

Table 1: Eyeblink8 videos and false triggers per minute (without → with the reject window).

Video	Glasses	Min	fps	Blinks	D1 FT/min	D3 FT/min
1	no	8.9	31.2	38	6.1 → 4.6	3.5 → 1.7
2	no	6.4	31.2	88	13.8 → 13.3	13.7 → 8.0
3	no	5.2	31.2	65	12.1 → 7.5	9.0 → 3.3
4	no	3.0	31.2	31	10.0 → 7.7	9.3 → 3.3
8	no	5.9	31.2	30	6.7 → 6.2	5.2 → 1.5
9	no	2.9	31.2	41	15.1 → 13.0	15.1 → 5.3
10	no	5.0	31.2	72	13.9 → 12.5	13.3 → 3.4
11	yes	2.7	31.2	43	16.9 → 16.6	14.7 → 6.6
Mean [95% CI] over videos					11.8 [9.2, 14.2] → 10.2 [7.5, 12.9]	10.5 [7.5, 13.2] → 4.1 [2.7, 5.7]

**Figure 3: Eyeblink8: measured closure against annotated blink length for matched blinks. The solid line is equality; the dashed line is the 200 ms reject window.**

middle of the lid movement. D1’s was a median 32 ms shorter, with a long tail of closures measured as much longer than the blink (-64 to 57 ms). Only 6% of these annotated blinks lasted under 200 ms, but 60% of D3’s measurements did, against 16% of D1’s. The reject window works on the measured duration, not the true one, so a window tuned for one signal does not transfer to the other.

What the false triggers would type. Table 2 shows how the segmenter would have used every closure the detectors found. With D1, natural blinks would have typed 6.6 dots, 1.4 dashes and 1.0 deletions per minute (pooled over all videos): a user would lose a character about once a minute without touching anything. With D3, the same blinks typed 3.4 dots and 0.3 dashes per minute and no deletions.

How long a window would be needed. Sweeping the reject window after the fact (Fig. 4, left; exploratory) shows that D3 reaches half its no-window rate at 200 ms, while D1 needs 320 ms. A 320 ms window would cover 71% of the annotated blinks, so with D1 the shortest deliberate closure would have to be longer than most natural blinks by a wide margin, which slows typing.

6 One Recorded Session (Exploratory)

Session. The author (able-bodied, no glasses) recorded one detection session of 9.6 minutes with a built-in 1080p laptop camera at

Table 2: Eyeblink8: what the segmenter would do with every detected closure, pooled over all videos: count (per minute). Every closure here is a natural blink or noise, so everything except “reject” is a false trigger.

Detector	reject	dot	dash	hold	backspace
D1	64 (1.6)	263 (6.6)	55 (1.4)	12 (0.3)	40 (1.0)
D3	222 (5.6)	137 (3.4)	13 (0.3)	2 (0.1)	0 (0.0)

Table 3: Session 1: calibration from the cue windows of the calibration block (20 closures). The calibrated threshold lies halfway between the open and closed medians; h is the hysteresis.

Signal	Open	Closed	Calibrated	h	Fixed
EAR, mean	0.315	0.042	0.178	0.027	0.21
Blink score, max	0.061	0.788	0.424	0.073	0.5
Blink score, left	0.052	0.725	0.388	0.067	0.5
Blink score, right	0.060	0.788	0.424	0.073	0.5

about 45 cm, in front lighting. A face was found in 99.7% of frames. The camera did not deliver 30 fps: it ran at 16.4 fps in the calibration block, and fell to about 10 fps in the last minute of the first reading block and the whole of the second (median per block: reading 15.1, cued 16.4, mixed 10.3 fps; Fig. 7). The video file shows the same pattern, so the drop came from the camera’s exposure, not from the app.

Everything in this section comes from one session of one person, without confidence intervals, and is exploratory. Closure durations here are *measured by the detectors*, not labelled, so they inherit each detector’s measurement error (Section 5).

Calibration. The calibrated detectors were fitted from the 20 cued closures of the calibration block, using the cue windows (Table 3), as the app does for a new user. The eyes-open EAR was 0.315 and the closed peaks 0.042, giving a calibrated threshold of 0.178 instead of the fixed 0.21. The blink score was 0.061 open and 0.788 closed, giving 0.424 instead of 0.5. Figure 6 shows the first 16 seconds of the cued block with both pairs of thresholds.

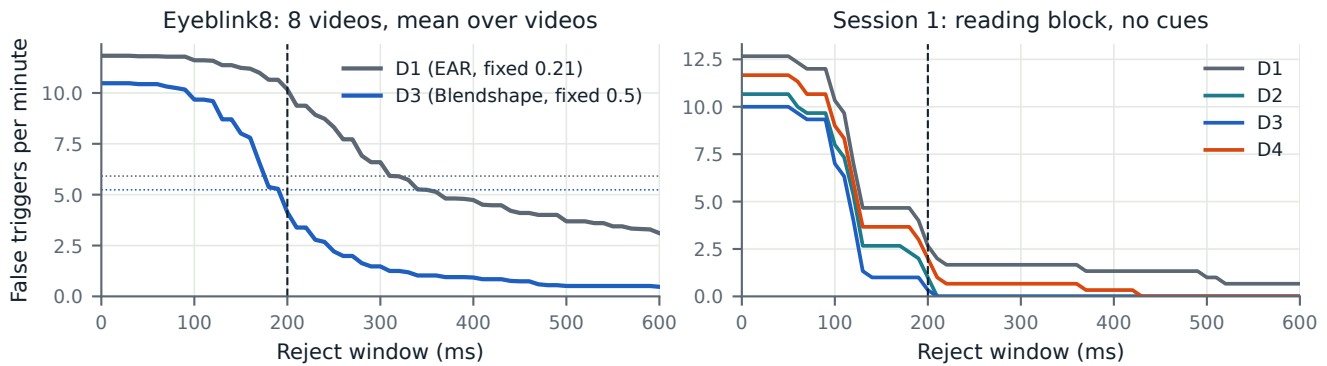


Figure 4: False triggers per minute against the reject window length (exploratory sweep, after seeing the data). Left: Eyeblink8, mean over videos; dotted lines mark half of each detector’s rate without a window. Right: session 1, reading block without cues. The dashed line is the default 200 ms.

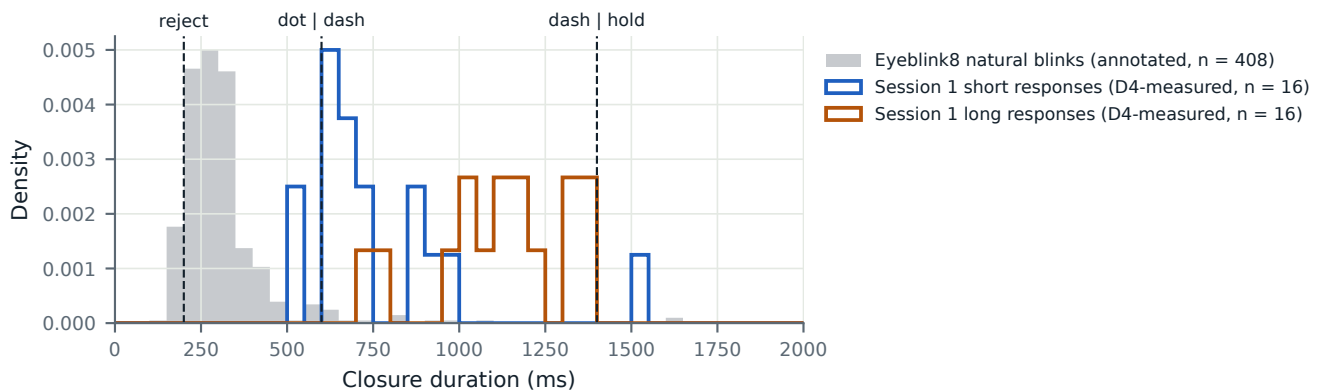


Figure 5: Closure durations against the default segmenter bands: annotated natural blinks of other people (Eyeblink8) and the author’s short and long responses in session 1, as measured by D4 (not labelled).

Table 4: Session 1: cued type against the response type checked by the author (counts).

Cue	Response checked by the author				
	short	long	wink-L	wink-R	none
short	24	1	0	0	1
long	0	26	0	0	0
wink-L	1	0	15	0	0
wink-R	0	0	1	15	0

Responses to cues. The author checked the response to every cue in the labelling tool (Table 4): 80 of 84 responses were of the cued type.

Reading without cues. The first reading block had no cues, so every closure detected in it counts as spontaneous. The both-eye detectors found 30 (D3) to 38 (D1) closures in three minutes, with median measured durations of 113 ms (D3) and 126 ms (D1). The reject window removed most of them: false triggers fell from 10.0

to 0.3 per minute with D3 and from 12.7 to 2.7 with D1 (Table 5). In this session most of the reduction happened between 100 and 200 ms (Fig. 4, right): at 100 ms, D3 still had 7.0 false triggers per minute, and at 300 ms D1 still had 1.7.

Short closures ran long. For each response in the cued and mixed blocks that the author checked as short or long (33 responses), the longest closure each detector found in the cue window (cue onset to 2.5 s after it) was classified by the default bands. With D4, 13 of 17 short responses fell in the dash band; only 2 were dots. The other detectors behaved similarly (Table 5). The measured short responses had a median of 684 ms and the long ones 1126 ms (Fig. 5). For this person the two kinds of closure were separated, but around a boundary well above the default 600 ms. This is what the planned threshold selection (preregistration, section 8) would have measured with labelled data.

Where the boundary would have been. Sweeping the dot-dash boundary over these responses (Fig. 8; exploratory) shows how much the default cost. With D4, 56% of the 32 responses were

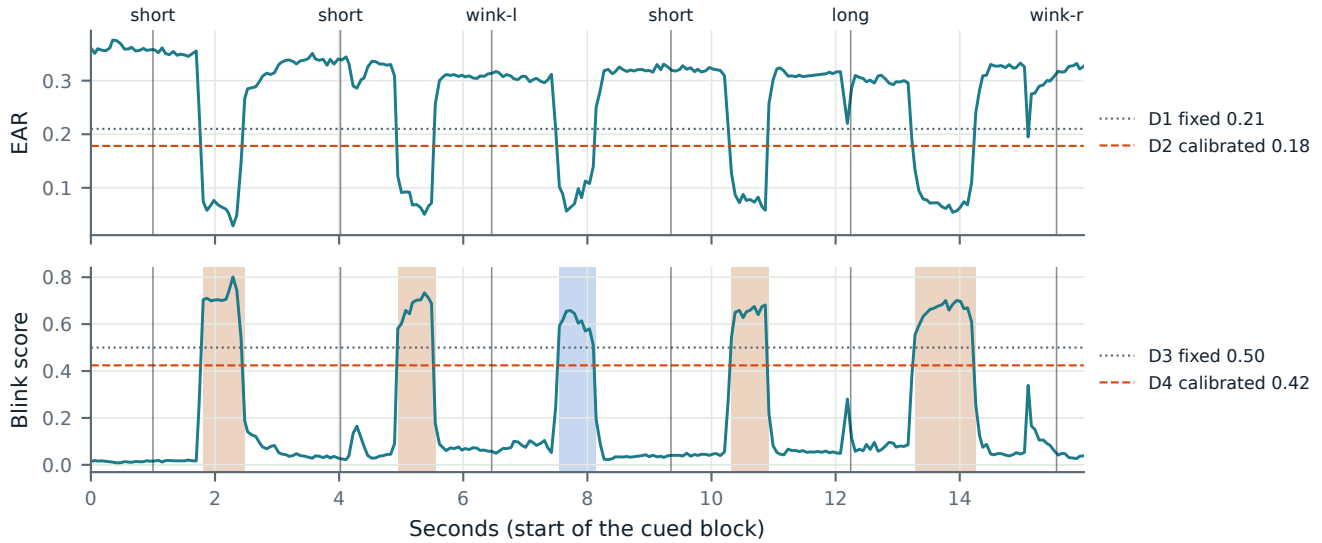


Figure 6: Session 1, start of the cued block: EAR (top) and blink score (bottom) with fixed and calibrated thresholds. Vertical lines are cues; shaded spans are D4's closures, coloured by segmenter band (blue dot, orange dash). Every cued short closure here falls in the dash band.

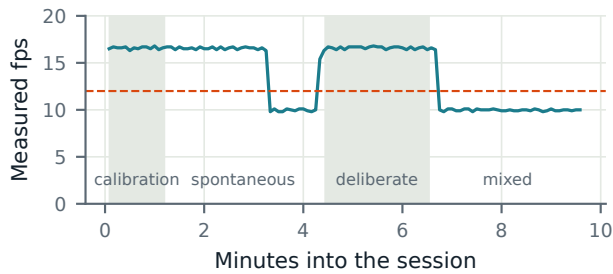


Figure 7: Session 1: measured frame rate by block. The dashed line is the preregistered 12 fps exclusion limit for a session's median.

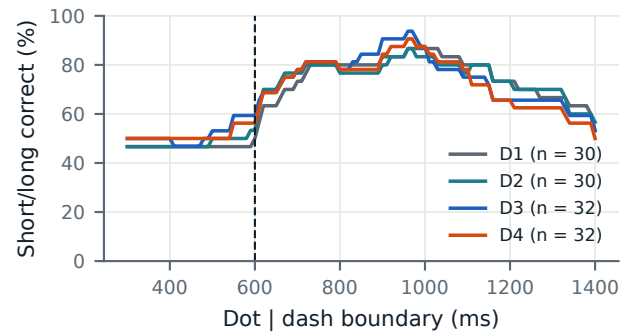


Figure 8: Session 1: short and long responses classified correctly against the dot-dash boundary (exploratory sweep over the same responses). The dashed line is the default 600 ms.

classified correctly at 600 ms; the best boundary, between 960 and 970 ms, reached 91% (D3: 59% and 94%). The boundary was chosen on the same data it is scored on, so the best accuracy is optimistic; the point is the size of the shift, not the exact value.

Winks closed both eyes. The wink detector (D5) found none of the author's winks. In every one of the 31 wink responses, the other eye's blendshape score was at or above 0.3 when the winking eye was most closed (median 0.66, against 0.74 for the winking eye). D5 requires it to stay below 0.3. Either the author's winks partly closed the other eye, or the blendshape scores of the two eyes move together. One session cannot tell these apart, but a fixed limit on the other eye did not work here.

7 Discussion

A duration reject window is an attractive defence against spontaneous blinks because it needs no model of the user. The public videos show that its effect depends on the signal it is applied to. The blendshape detector measured natural blinks as short enough that most fell inside a 200 ms window. The fixed EAR threshold measured the same blinks as longer, so the same window removed few. A reject window therefore has to be chosen for the signal and threshold in use, not carried over from published blink durations alone.

The single session points to the other half of the problem: deliberate closures have to be reliably longer than the window, and

Table 5: Session 1 by detector: false triggers per minute while reading (without → with the reject window); cue responses whose longest detected closure fell in the cued band; and median measured duration of short and long responses.

ID	Detector	Reading FT/min	Correct band	Short (ms)	Long (ms)
D1	EAR, fixed 0.21	12.7 → 2.7	13/33	702	1242
D2	EAR, calibrated	10.7 → 1.0	15/33	635	1153
D3	Blendshape, fixed 0.5	10.0 → 0.3	18/33	664	1118
D4	Blendshape, calibrated	11.7 → 2.0	17/33	684	1126
D5	Wink (blendshape, calibrated)	0.3 → 0.3	0/31	–	–

the dot–dash boundary has to fit the person. The author’s cued short closures, measured from the camera, mostly exceeded the default dot limit. With the defaults, most dots would have been typed as dashes; a boundary near 960 ms would have separated most of them. A per-user timing calibration, analogous to the threshold calibration, is the natural next step, and the planned study would have tested one.

Frame rate matters for both. At 10 fps a frame lasts 100 ms, half the reject window, so durations near the window’s edge cannot be measured reliably. The camera in this session dropped to about 10 fps for long stretches, apparently through its automatic exposure; a deployed system would need to detect and report this.

8 Limitations

The preregistered study was not completed; none of its confirmatory hypotheses was tested on the author’s data, and the typing study was not run. The public-video analysis is not blind and covers four people in eight videos. The session results come from one session of one person, who also designed the system, knew the hypotheses and checked the response types. Session durations are measured by detectors, not labelled, and the response types were checked against placeholder windows, not exact closure times. The camera ran below the planned frame rate. No one with a motor impairment used the system, and no clinical claim is made.

9 Ethics and Data

No ethics review was sought. The only participant is the author, and the public videos are used under their published licence (GPL-3.0). The app processes video on the device and logs only per-frame features; no faces, video or names are published. The session’s log and response types are released under CC BY 4.0 in `data/public/`, with a datasheet (`data/DATASHEET.md`).

10 Conclusion

MorsEye shows that blink-based Morse typing can run entirely in a browser on a laptop webcam, but also how much depends on timing. On other people’s natural blinks, a 200 ms reject window more than halved false triggers with a blendshape detector but barely changed them with a fixed EAR threshold. In one exploratory session, the default dot limit was too short for the author’s deliberate closures. Choosing the reject window and the dot–dash boundary per signal and per person, and testing that choice in the study this project planned, is the work that remains.

References

- [1] Tomas Drutarovsky and Andrej Fogelton. 2014. Eye Blink Detection Using Variance of Motion Vectors. In *Computer Vision – ECCV 2014 Workshops (Lecture Notes in Computer Science, Vol. 8927)*. Springer. doi:10.1007/978-3-319-16199-0_31
- [2] Andrej Fogelton and Wanda Benesova. 2016. Eye Blink Detection Based on Motion Vectors Analysis. *Computer Vision and Image Understanding* 148 (2016), 23–33.
- [3] Google. [n.d.]. MediaPipe Face Landmarker. https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker. Version 0.10.21.
- [4] K. Grauman, M. Betke, J. Lombardi, J. Gips, and G. Bradski. 2003. Communication via eye blinks and eyebrow raises: video-based human-computer interfaces. *Universal Access in the Information Society* (2003).
- [5] P. Majaranta and K.-J. R  ih  . 2002. Twenty Years of Eye Typing: Systems and Design Issues. In *Proceedings of the Symposium on Eye Tracking Research & Applications (ETRA)*.
- [6] T. Soukupov   and J.   ech. 2016. Real-Time Eye Blink Detection using Facial Landmarks. In *Computer Vision Winter Workshop (CVWW)*.