Title

Real-Time Estimation of Listener Comprehension Using Blink Synchronization

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Saito, Koki

Enomoto, Makoto

Zenio, Haruhito

et al.

Publication Date

2026

Real-Time Estimation of Listener Comprehension Using Blink Synchronization

Koki Saito (koki.saito@biprogy.com)[1], Makoto Enomoto[1], Haruhito Zenio[1], Takayuki Hoshino[1] & Eiji Takahashi[1]

[1]BIPROGY Inc.

Abstract

Blink synchronization is closely associated with cognitive state and comprehension during communication. This study developed a real-time model to estimate listener comprehension using blink synchronization. Data were collected from 22 participants who watched a 3-minute explanatory video. An estimation model was trained on the full dataset, with blink synchronization as a key feature. The model discriminated between comprehension levels, with relatively balanced performance (accuracy = 0.727, F1-score = 0.750, and ROC-AUC = 0.726). Real-time comprehension probabilities were then calculated using a 60-second sliding window updated every second. These estimates generally aligned with overall comprehension labels and suggest group-level stability, with limited late-stage deviations among some low-comprehension participants. Overall, the findings highlight the potential feasibility of blink synchronization as a dynamic indicator of listener understanding and contribute to the development of tools for real-time communication support.

Keywords: blink synchronization; listener comprehension; communication; real-time estimation; cognitive state; physiological signals

Introduction

Accurately estimating others' internal states, such as their level of comprehension, is essential for effective communication. People tend to develop greater liking, self–other overlap, and empathy toward individuals who demonstrate understanding and perspective-taking (Goldstein et al., 2014). Feeling understood by others is also fundamental to the formation of positive interpersonal relationships (Livingstone, 2023).

Although people often attempt to infer others' internal states from verbal and nonverbal cues (Astington, 1993), these inferences are frequently inaccurate (O'Sullivan, 2005). For instance, emotional states are commonly inferred from facial expressions, yet such interpretations vary across cultures and contexts, making accurate recognition difficult (Barrett et al., 2019; Sato et al., 2019). Moreover, individuals can deliberately conceal their true intentions (Bonanno et al., 2004; Gross, 2002), further complicating accurate interpretation.

Nonverbal behavior plays a critical role in estimating another person's internal state (Mann et al., 2013; Vrij et al., 2019), particularly from the speaker's perspective. During communication, speakers cannot directly access a listener's internal state through verbal feedback alone and must therefore rely on nonverbal signals to infer comprehension. For instance, salespeople often use customers' nonverbal cues to infer preferences (Honda et al., 2016). Importantly, certain physiological signals, such as heart rate and eye movements, are difficult to consciously control (Cruz et al., 2011), making them difficult to disguise and potentially valuable for assessing internal states.

One specific nonverbal behavior is blink synchronization, which reflects a listener's engagement and comprehension during interpersonal interactions. Prior research has shown that a listener's blinks can spontaneously synchronize with those of a salesperson when the listener is interested (Nakano & Miyazaki, 2019). More broadly, inter-individual neural synchronization is known to occur during social interactions (Hari & Kujala, 2009; Saito et al., 2010). However, neural synchronization requires specialized equipment, such as functional magnetic resonance imaging or near-infrared spectroscopy, making it impractical for everyday communication. Conversely, spontaneous blinks can be captured using simple devices such as webcams and are not under conscious control. Therefore, they offer a practical and unobtrusive means of assessing comprehension. Although blink rate has been associated with internal state (Cho et al., 2000; Karson et al., 1981), it is highly variable across individuals and contexts (Nakano et al., 2015). Consequently, blink synchronization provides a more stable and informative signal than blink rate alone for estimating internal states (Nakano & Miyazaki, 2019).

Blink synchronization reflects unconscious information sharing between individuals (Nakano & Kitazawa, 2010). Individuals spontaneously synchronize their blinks at both explicit and implicit breakpoints in ongoing information streams (Nakano et al., 2009; Nakano & Kitazawa, 2010). For instance, when watching the same video, viewers tend to blink synchronously at implicit narrative boundaries (Nakano et al., 2009). Individuals can blink to process information received at these breakpoints. The default mode network (DMN) is activated by blinking, which reflects information processing (Nakano et al., 2013). In communication, listeners synchronize their blinks with a delay of a few hundred milliseconds at breakpoints, such as the end and during pauses of a speaker's speech (Nakano & Kitazawa, 2010). Additionally, blink synchronization delay is considered a result of the comprehension time lag (Nakano & Miyazaki, 2019). Therefore, listeners who receive information from speakers and comprehend it well are likely to synchronize their blinks with the speakers.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

A prior study reported that listeners with high comprehension unconsciously synchronized their blinks with the speaker's blinks, with a delay of a few hundred milliseconds (K. Saito et al., 2025). A relationship between comprehension and blink synchronization emerges in situations requiring comprehension. Listener comprehension levels can be estimated using blink synchronization. Visualizing individuals' level of comprehension can contribute to smoother communication by assisting individuals with understanding how well others comprehend it.

Real-time implementation is desirable for monitoring comprehension during communication. Most studies have examined blink synchronization over several minutes, with only minimal studies evaluating real-time tracking, which can be enabled using blink synchronization with shorter windows. However, the prior dataset did not include a real-time assessment of comprehension, which was difficult to evaluate. Therefore, this study focuses on the feasibility of real-time comprehension estimation and examines the stability and consistency of these estimates relative to overall comprehension levels.

Based on this rationale, two hypotheses were proposed.

H1 (classification): Blink synchronization with the speaker can classify differences in listener comprehension at above-chance levels.

H2 (real-time tracking): Comprehension time series derived from shorter windows of blink synchronization remain stable within participants over time and are consistent with their overall comprehension labels.

This study contributes an estimation framework that leverages blink synchronization to classify listener comprehension, as well as a real-time implementation that generates short-window, time series probability trajectories. Collectively, these advances propose an unobtrusive approach for monitoring comprehension, suggest stability and label alignment at the group level, and lay the foundation for adaptive, feedback-driven communication systems that do not require specialized hardware.

Method

In this study, blink synchronization was operationalized to develop both an estimation model (H1) and real-time estimation framework (H2) using a previously collected dataset. The dataset reported by K. Saito et al. (2025) was used to estimate listeners' levels of comprehension based on blink synchronization. It comprises 3-minute video recordings of participants viewing an explanatory video, along with a post-viewing questionnaire assessing comprehension on a five-point scale. Blink synchronization was calculated for each recording and used as the primary feature for subsequent analyses.

Dataset

This dataset comprised video recordings of 22 Japanese adults (eight men and 14 women; four in their 20s, eight in their 30s, seven in their 40s, and three in their 50s), each of whom viewed one of three approximately 3-minute video clips explaining a research topic. The same speaker delivered all three videos using the same logical structure, with minimal differences in word count, prosodic pitch, and blink patterns, thereby reducing stimulus-driven variability across videos. Participants viewed the videos in their usual working environments, with no restrictions imposed. After viewing, participants completed a questionnaire assessing comprehension on a five-point scale. The dataset analyzed in this study was collected with prior ethics approval and informed consent from the original project [committee name withheld for review]. This study constitutes a secondary analysis conducted by the original research team within the scope of the approved protocol.

Detection of Blink Onset

Blinks were detected using personalized eye aspect ratios (EARs) derived from three-dimensional (3D) facial landmarks (Saito, 2024) to account for individual differences in eye shape and blinking behavior (Dewi et al., 2022). Facial 3D landmarks were extracted using Google MediaPipe (Lugaresi et al., 2019). The eye region was defined by 16 landmarks (Figure 1). EAR was computed using Equation (1) (Kraft, Hartmann, & Bieber, 2023) based on these landmarks. When the eyes are closed, EAR approaches zero; when open, it approaches a stable constant.

$$EAR \stackrel{\text{def}}{=} \frac{\|P3-P13\|_3 + \|P4-P12\|_3 + \|P5-P11\|_3}{3\cdot\|P0-P8\|_3} \quad (1)$$

EAR with eyes closed was calculated using Equation (2) and EAR with eyes open was calculated using Equation (3). The personalized EAR threshold is calculated using Equation (4). If the EAR was less than or equal to the personalized EAR threshold, it was considered a blink (Figure 2).

$$EAR_{Closed} \stackrel{\text{def}}{=} \frac{min(\|P3-P13\|_3)+min(\|P4-P12\|_3)+min(\|P5-P11\|_3)}{3\cdot max(\|P0-P8\|_3)} \quad (2)$$

$$EAR_{Open} \stackrel{\text{def}}{=} \frac{mean(\|P3-P13\|_3)+mean(\|P4-P12\|_3)+mean(\|P5-P11\|_3)}{3\cdot min(\|P0-P8\|_3)} \quad (3)$$

$$Personalized\ EAR_{Threshold} \stackrel{\text{def}}{=} \left(EAR_{Open} - EAR_{Closed}\right)/2 \quad (4)$$

The peaks were then calculated and the peak that appeared immediately before the blink was defined as blink onset (Figure 2).

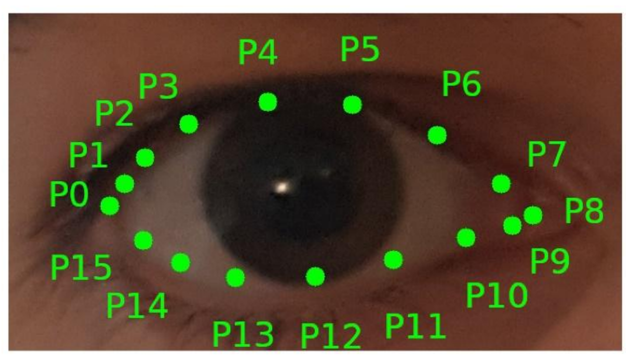

Figure 1: Sixteen landmarks around the eye (Kraft et al., 2023).

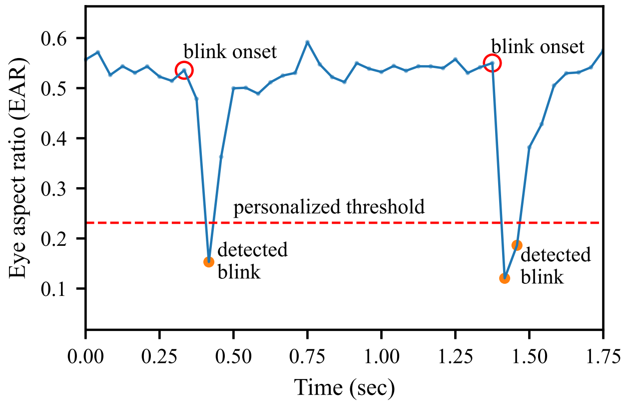

Figure 2: Example time series data of the eye aspect ratio.

Detection of Blink Synchronization

Blink synchronization was calculated using the method proposed by Nakano and Miyazaki (2019). For each dyad, individuals were designated as the reference blinker (speaker) and test blinker (listener). Blink onset asynchrony was calculated as the difference between listener and speaker blink onsets; positive values indicated that listener blinks followed speaker blinks. Histograms of blink onset asynchrony were constructed with a central window of $\pm$ 0.3 seconds around zero and a total range of $\pm$ 2.4 seconds (bin width: 0.6 seconds). Z-score normalization was performed using 1,000 surrogate datasets by randomly shuffling the inter-blink intervals of the test blinker while preserving their temporal structure. A positive z value indicates that the observed frequency exceeds chance-level probability. Following prior work, the z-value in the 0.6 seconds bin (0.3–0.9 seconds) was used as the blink synchronization score, as listeners are known to synchronize their blinks with a delay of a few hundred milliseconds relative to the speaker (Nakano & Kitazawa, 2010; Nakano & Miyazaki, 2019; K. Saito et al., 2025).

Estimation Model of Listener Comprehension

Participants were divided into two groups based on their self-reported comprehension scores. Given the cultural tendency among Japanese individuals to avoid strong positive self-enhancement (De Jong et al., 2008; Heine et al., 1999), along with the observed score distribution (no participant scored five points, mean = 2.95, median = 3), participants who scored two or below were assigned to the "Not comprehend" group (nine participants), and those scoring three or above were assigned to the "Comprehend" group (13 participants).

Logistic regression was applied to classify between the "Comprehend" and "Not comprehend" groups. Blink synchronization scores at 0.6 seconds (0.3–0.9 seconds) and the sum of band-integrated probability mass across −0.3–0.3 seconds, 0.3–0.9 seconds, and 0.9–1.5 seconds were used as the explanatory variables. This selection followed a prior study reporting a significant difference between the "Comprehend" and "Not comprehend" groups (Figure 3, $t_{20} = 2.01, p < 0.05$) at 0.6 seconds (0.3–0.9 seconds) (K. Saito et al., 2025). Band integration was included because integrating over bands reduces the sensitivity to bin alignment and small timing jitter compared to a single-bin variable and captures a broader delay structure. The model employed with L1 regularization with inverse regularization strength C = 100 and the liblinear solver, with class weights set to 'balanced' to mitigate class imbalance. The classification threshold was set at the default value of 0.5 (no threshold tuning), because only the probability was used in the subsequent real-time estimation. Logistic regression was implemented using Scikit-learn (1.6.1) in Python.

Model performance was evaluated using leave-one-out cross-validation (LOOCV), in which the model was trained on 21 participants and tested on the remaining participant. This procedure was repeated for each participant. To assess generalizability across videos, a leave-one-video-out (LOVO) scheme was also employed, in which the model was trained on participants from two videos and tested on participants from the third.

Accuracy, precision, recall, F1-score, balanced accuracy, and ROC-AUC were used as evaluation metrics; for the LOVO setting, only ROC-AUC was reported. "Comprehension" was defined as a positive class that prioritizes the detection of positive feedback that a listener exhibits, which is critical for facilitating effective communication. Out-of-fold predicted probabilities from LOOCV were used to compute ROC-AUC values. Confidence intervals were estimated using participant-level bootstrapping (5,000 resamples; percentile/BCa method), and significance relative to chance (AUC = 0.5) was assessed using permutation tests (10,000 label shuffles).

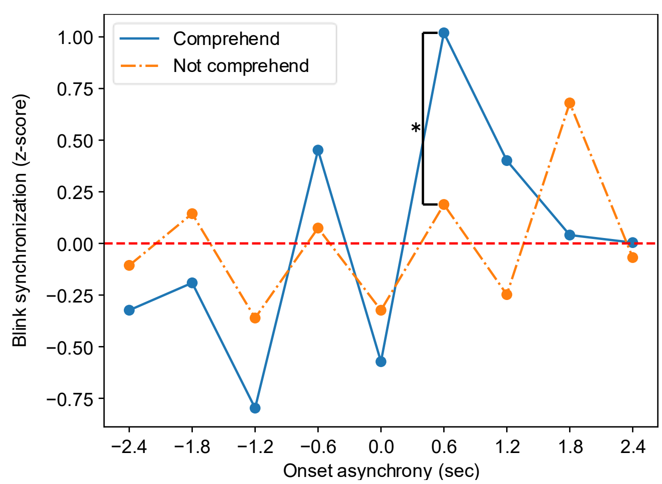

Figure 3: Blink synchronization between the speaker and all listeners as a function of comprehension level ($*$: $p < .05$).

Real-Time Estimation of Listener Comprehension

For real-time estimation, blink synchronization was calculated using a trailing 60-second sliding window updated every second. This approach produces a time series that reflects dynamic changes in synchronization. Given that individuals blink approximately once every 3 seconds on average (Bentivoglio et al., 1997), a sufficient number of blinks are required to reliably calculate synchronization within the 60 seconds window. Nakano and Miyazaki (2019) calculated blink synchronization using a 30 seconds window, extending the window to 60 seconds ensures a more stable calculation by incorporating a larger number of blinks.

To validate the reliability of blink synchronization detection and account for individual differences, surrogate data were generated using the entire 3-minute dataset. This approach addresses the potential limitations of the 60 seconds window, which may not capture sufficient data to reliably measure individual characteristics, thereby ensuring robust synchronization calculations. In online settings, a calibration period is used and surrogate data are generated based on past data to avoid the use of future information.

Logistic regression probability was used as a continuous indicator of the level of comprehension, because it provides a continuous measure of the likelihood of comprehension. The probability is defined as:

$$p(x) \stackrel{\text{def}}{=} \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\beta_2 x_2)}}, \tag{5}$$

where β_0 is the intercept, while β_1 and β_2 are the coefficients; x_1 is the blink synchronization score at 0.6 seconds (0.3–0.9 seconds), and x_2 is the sum of band-integrated probability mass over −0.3–0.3 seconds, 0.3–0.9 seconds, and 0.9–1.5 seconds. Probabilities were calculated using the same LOOCV scheme as the classification model. To provide real-time feedback for real-time estimation in the future, it is deemed appropriate to represent the information as numerical data that change dynamically over time. If the probability is above 0.5, the participant is defined as "Comprehend."

To test the stability and consistency of the real-time estimates, a one-tailed t-test, consistency rate (defined as the proportion of windows whose predicted label matches the participant's overall label), and a dip test (Hartigan & Hartigan, 1985) were conducted. One-tailed t-tests were used to test the differences between the probabilities of the two groups. The consistency rate was used to analyze the consistency over time. The dip test (Hartigan & Hartigan, 1985) was used to determine whether the probability distribution over time was unimodal. Statistical analyses were performed using Python version 3.12.5. Inferential tests were conducted on participant-level summaries (e.g., per-participant means, proportions, and counts) rather than on window-level data to avoid pseudo-replication owing to overlapping windows and within-participant autocorrelation.

Results

The classifier performance for overall comprehension (H1) was first reported, followed by participant-level stability and consistency analyses for real-time probabilities (H2).

Estimation Model of Listener Comprehension

Figure 4 presents the confusion matrix of the logistic regression classifier, and Table 1 summarizes its performance. Precision (0.818) was the highest among the four primary metrics, indicating relatively few false positives and reliable identification of the "Comprehend" group. Accuracy (0.727) exceeded the majority-class baseline (0.591, calculated as 13/22), demonstrating that the model successfully distinguished between the "Comprehend" and "Not comprehend" groups. Recall (0.692), however, was lower than precision, indicating that some "Comprehend" cases were missed. Despite this asymmetry, the F1 score (0.750), which balances precision and recall, and the balanced accuracy (0.735) indicate that the model achieved relatively well-balanced overall performance. The ROC-AUC (0.726) further indicated above-chance discrimination with a moderate effect size, albeit with some uncertainty, likely because of the small sample size. Collectively, these metrics suggest relatively robust and balanced classification performance, even though some true "Comprehend" cases were missed.

The LOVO evaluation yielded a slightly lower ROC-AUC (0.692) than the participant-level LOOCV (ROC-AUC = 0.726), suggesting that stimulus-related differences among the three videos had only a limited impact on performance.

Consistent with H1, the model exceeded the majority-class baseline and achieved balanced classification performance. These findings suggest that blink synchronization may provide a sufficient and informative signal for distinguishing between the two comprehension groups within this dataset.

Table 1: Performance of the estimation model ("Comprehend" as the positive class).

Metric	Value	95% CI (BCa)	p_perm (AUC > 0.5)
Accuracy	0.727	[0.46-0.86]	-
Precision	0.818	[0.56-1.00]	-
Recall	0.692	[0.31-0.85]	-
F1-score	0.750	[0.50-0.89]	-
Balanced ACC	0.735	[0.51-0.89]	-
ROC-AUC	0.726	[0.43-0.92]	0.026
ROC-AUC (LOVO)	0.692	[0.39-0.88]	0.049

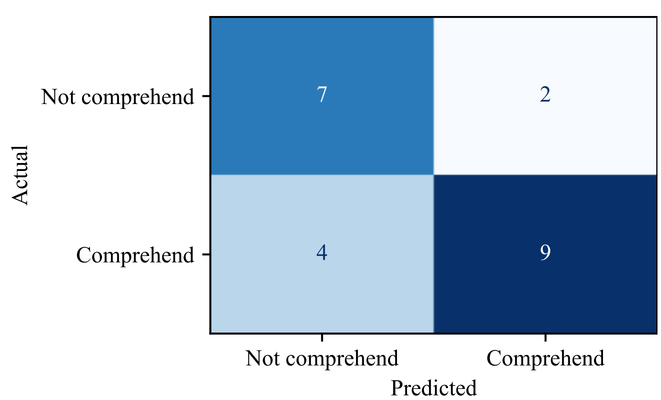

Figure 4: Confusion matrix.

Real-Time Estimation of Listener Comprehension

Figure 5 exhibits the time series of mean estimated comprehension probabilities for two groups ("Comprehend" and "Not comprehend"). In the "Comprehend" group, probabilities generally remained at or above 0.5, although a temporary decrease was observed between approximately 90 and 120 seconds. There was no significant difference between the mean probability (0.552) in the "Comprehend" group and that in the "Not comprehend" group (0.439) (Figure 6:$t_{20} = 1.05, p = 0.153$). The mean consistency rate in the "Comprehend" group was 0.583 and exceeded the chance level of 0.5 (Figure 7).

In the "Not comprehend" group, the probability tended to remain below 0.5 until approximately 130 seconds, after which it shifted to values relatively at 0.5 or above. This shift suggests that some participants who rated lower levels of comprehension likely synchronized their blinks with the speaker, because the probability was calculated based on the blink synchronization score (Equation (5)). The mean consistency rate in the "Not comprehend" group was 0.609 and exceeded the chance level of 0.5 (Figure 7). These findings suggest that the model exhibited relative stability and consistency, although there were deviations from the overall comprehension labels in both groups.

Figure 8 exhibits a violin plot of the estimated probability divided into two time ranges: 60-130 seconds and 130-180 seconds. The comprehension group's probabilities tended to remain at 0.5 or higher in both ranges. However, the "Not comprehend" group had bimodal distribution in the time 130-180 seconds range ($D = 0.167, p < 0.01$), indicating the presence of the two subgroups. The mean consistency rate in the "Not comprehend" group was 0.726 within 60-130 seconds and 0.444 within 130-180 seconds. These findings imply that some participants in the "Not comprehend" group synchronized their blinks with the speaker during the latter part of the video, despite rating their overall comprehension as low. Consequently, the probabilities in the "Not comprehend" group tended to exceed 0.5 in the latter part of the time series. Some participants' estimated comprehension states may have changed over time rather than remaining constant throughout the 3-minute video. Therefore, stability and consistency were confirmed within 60–130 seconds but not within 130–180 seconds.

These findings partially support H2 and suggest the existence of time-varying comprehension. The real-time comprehension probability exhibited relative stability and consistency in both groups, except for some deviations: "Comprehend" within 90-120 seconds and "Not comprehend" within 130-180 seconds.

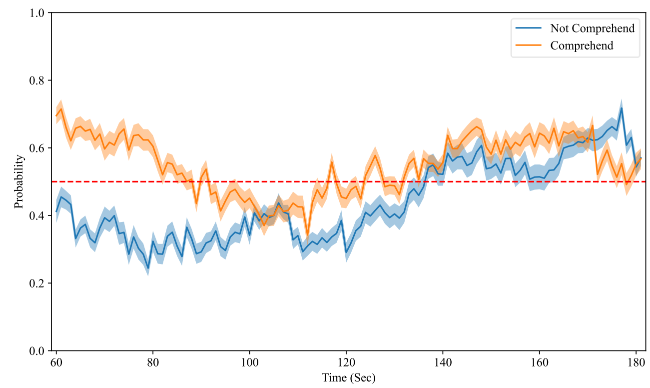

Figure 5: Time series of the mean estimated probability as a function of comprehension. Error bars represent standard errors, and the red horizontal line indicates the decision threshold (0.5).

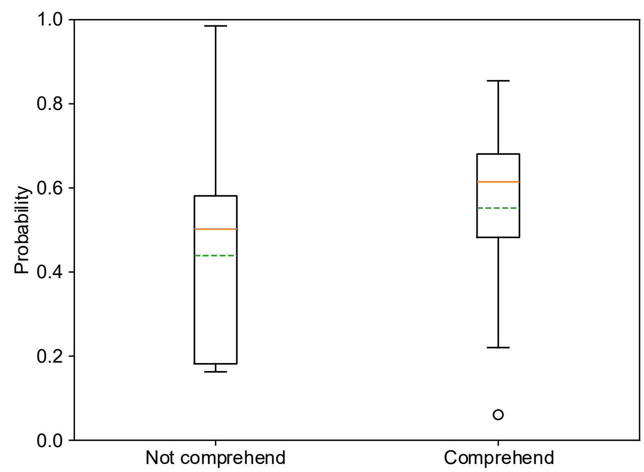

Figure 6: Box plot of mean estimated probability by differences in comprehension.

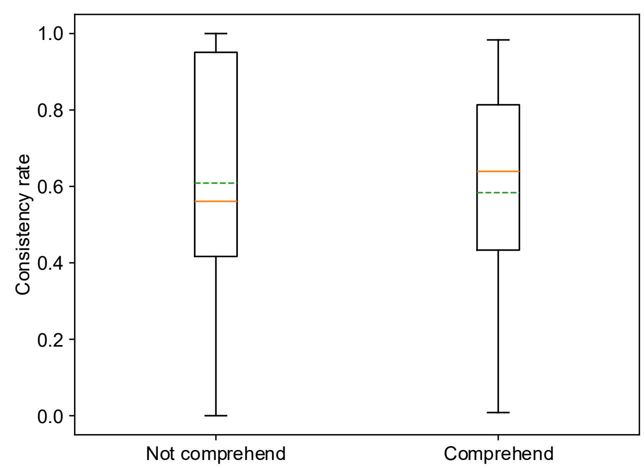

Figure 7: Box plot of consistency rate by differences in comprehension.

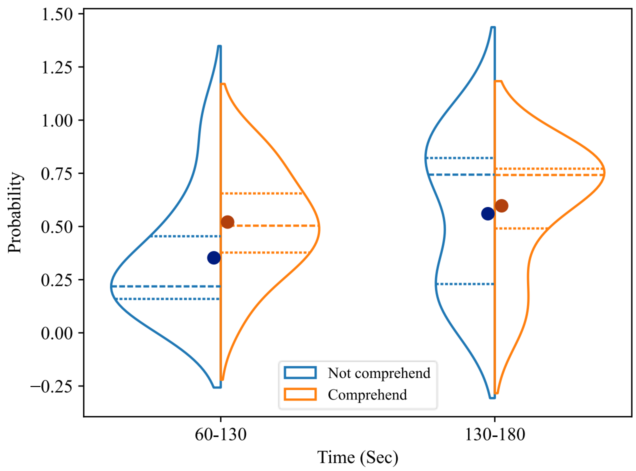

Figure 8： Violin plot of mean estimated probability by time range (60-130 seconds vs 130-180 seconds) as a function of comprehension.

Discussion

This study adopted a stepwise validation approach. First, consistent with H1, we developed an estimation model of listener comprehension based on blink synchronization, building on prior work (K. Saito et al., 2025). The model achieved balanced classification performance, supporting the feasibility of using blink synchronization as an index of comprehension. Second, H2 was partially supported. We implemented a real-time estimation framework using a 60-second sliding window. The resulting time series probabilities were generally aligned with overall comprehension labels and showed relative temporal stability. However, dynamic changes were observed among some participants in the Not Comprehend group, preventing full support for stability across both groups. Importantly, the real-time model was able to capture these temporal variations, highlighting its potential feasibility for real-world communication. Given the limited prior research on real-time comprehension estimation, this approach represents a meaningful step toward adaptive communication systems.

Our findings suggest that estimated comprehension may change dynamically even during short (3-minute) video viewing, influencing both blink synchronization and model predictions. The bimodal distribution observed in the Not Comprehend group (Figure 8) suggests that some participants who reported low overall comprehension nevertheless exhibited high estimated probabilities in the latter part of the video. This may reflect delayed or partial understanding, particularly during the summary phase (approximately 130–140 seconds), when information load decreases. Narrative comprehension involves integrative processing and frequent engagement of the default mode network (DMN; Song et al., 2021). Because the DMN is also activated by blinking (Nakano et al., 2013), this integrative process may have increased blink synchronization toward the end of the videos. However, such late-stage understanding may not have been sufficient to influence participants' overall self-ratings, likely contributing to reduced recall in the classification model (Table 1).

Temporal variability may also explain misclassifications in the Comprehend group. Some participants with high self-reported comprehension did not synchronize their blinks consistently throughout the full video (Figure 4). Because the questionnaire captured only overall comprehension, it failed to reflect temporal fluctuations. Overall comprehension does not necessarily depend on the duration of perceived understanding, as insight can occur abruptly (e.g., through an "aha" experience). In contrast, blink synchronization reflects sustained engagement. Consequently, models based on aggregated data may fail to capture brief but meaningful cognitive events.

Blink synchronization likely reflects internal cognitive processing, as blinking activates the DMN (Nakano et al., 2013), which supports self-referential thought, memory, and language (Menon, 2023; Raichle, 2015). Thus, blink synchronization may index listeners' cognitive effort during comprehension. However, its functional meaning is context dependent. In explanatory contexts, it appears to reflect understanding, whereas in low-demand contexts it may reflect interest or engagement (Nakano & Miyazaki, 2019). Although blink synchronization can reflect multiple processes, these processes are generally positive and essential for smooth communication.

The proposed real-time model offers practical advantages because it captures dynamic changes in comprehension rather than relying on static, post hoc assessments. In real-world interactions, comprehension fluctuates, and feedback based solely on overall comprehension may be insufficient. Real-time feedback could enable speakers—such as educators, salespeople, and healthcare professionals—to adapt their communication strategies dynamically. Because blink synchronization can be measured using a standard webcam, the model is technically feasible for everyday use.

Several limitations should be noted. The sample size was small ($N = 22$) and culturally homogeneous, limiting generalizability. Comprehension was assessed only via self-report, which does not capture objective or time-resolved changes in understanding. Additionally, viewing conditions were uncontrolled, introducing potential noise. However, these unconstrained conditions may enhance ecological validity, and the present results should be interpreted as a proof-of-concept given the small sample size and reliance on a single self-report measure of comprehension.

Future research will evaluate the real-world utility of this model, including whether real-time comprehension feedback improves communication outcomes. We plan to test the system in dyadic and collaborative tasks, which will provide insight into its potential applications in education, teamwork, and interpersonal communication.

References

Astington, J. W. (1993). *The child's discovery of the mind*. Cambridge, MA: Harvard University Press.

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*(1), 1–68. https://doi.org/10.1177/1529100619832930

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders*, *12*(6), 1028–1034. https://doi.org/10.1002/mds.870120629

Bonanno, G. A., Papa, A., Lalande, K., Westphal, M., & Coifman, K. (2004). The importance of being flexible: The ability to both enhance and suppress emotional expression predicts long-term adjustment. *Psychological Science*, *15*(7), 482–487. https://doi.org/10.1111/j.0956-7976.2004.00705.x

Cho, P., Sheng, C., Chan, C., Lee, R., & Tam, J. (2000). Baseline blink rates and the effect of visual task difficulty and position of gaze. *Current Eye Research*, *20*(1), 64–70. https://doi.org/10.1076/0271-3683(200001)2011-HFT064

Cruz, A. A. V., Garcia, D. M., Pinto, C. T., & Cechetti, S. P. (2011). Spontaneous eyeblink activity. *The Ocular Surface*, *9*(1), 29–41. https://doi.org/10.1016/S1542-0124(11)70007-6

De Jong, Martijn G, Steenkamp, Jan-Benedict E M, Fox, Jean-Paul, & Baumgartner, Hans. (2008). Using item response theory to measure extreme response style in marketing research: a global investigation. Journal of Marketing Research, 45(1), 104–115. https://doi.org/10.1509/jmkr.45.1.104

Dewi, C., Chen, R.-C., Jiang, X., & Yu, H. (2022). Adjusting eye aspect ratio for strong eye blink detection based on facial landmarks. *PeerJ Computer Science*, *8*, e943. https://doi.org/10.7717/peerj-cs.943

Goldstein, N. J., Vezich, I. S., & Shapiro, J. R. (2014). Perceived perspective taking: When others walk in our shoes. *Journal of Personality and Social Psychology*, *106*(6), 941–960. https://doi.org/10.1037/a0036395

Gross, J. J. (2002). Emotion regulation: Affective, cognitive, and social consequences. *Psychophysiology*, *39*(3), 281–291. https://doi.org/ 10.1017/S0048577201393198

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: From concepts to brain imaging. *Physiological Reviews*, *89*(2), 453–479. https://doi.org/10.1152/physrev.00041.2007

Hartigan, J. A., & Hartigan, P. M. (1985). The Dip Test of Unimodality. The Annals of Statistics, 13(1), 70–84. http://www.jstor.org/stable/2241144

Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? Psychological Review, 106(4), 766.

Honda, H., Hisamatsu, R., Ohmoto, Y., & Ueda, K. (2016). Interaction in a natural environment: Estimation of customer's preference based on nonverbal behaviors. *Proceedings of the Fourth International Conference on Human Agent Interaction* (pp. 93–96). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/2974804.2980512

Karson, C. N., Berman, K. F., Donnelly, E. F., Mendelson, W. B., Kleinman, J. E., & Wyatt, R. J. (1981). Speaking, thinking, and blinking. *Psychiatry Research*, *5*(3), 243–246. https://doi.org/10.1016/0165-1781(81)90070-6

Kraft, D., Hartmann, F., & Bieber, G. (2023). Camera-based blink detection using 3D-landmarks. In *Proceedings of the 7th International Workshop on Sensor-Based Activity Recognition and Artificial Intelligence* (Article no. 13, pp. 1–7). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/3558884.3558890

Livingstone, A. G. (2023). Felt understanding in intergroup relations. *Current Opinion in Psychology*, *51*, 101587. https://doi.org/10.1016/j.copsyc.2023.101587

Lugaresi, C., Tang, J., Nash, H., McClanahan, C., Uboweja, E., Hays, M., Zhang, F., Chang, C.-L., Yong, M. G., Lee, J., Chang, W.-T., Hua, W., Georg, M., & Grundmann, M. (2019). MediaPipe: A framework for building perception pipelines. *ArXiv*, abs/1906.08172. https://doi.org/10.48550/arXiv.1906.08172

Mann, S., Ewens, S., Shaw, D., Vrij, A., Leal, S., & Hillman, J. (2013). Lying eyes: Why liars seek deliberate eye contact. *Psychiatry, Psychology and Law*, *20*(3), 452–461. https://doi.org/10.1080/13218719.2013.791218

Menon, V. (2023). 20 years of the default mode network: A review and synthesis. Neuron, 111(16), 2469–2487. https://doi.org/10.1016/j.neuron.2023.04.023

Nakano, T., Kato, M., Morito, Y., Itoi, S., & Kitazawa, S. (2013). Blink-related momentary activation of the default mode network while viewing videos. Proceedings of the National Academy of Sciences, 110(2), 702–706. https://doi.org/10.1073/pnas.1214804110

Nakano, T., & Kitazawa, S. (2010). Eyeblink entrainment at breakpoints of speech. *Experimental Brain Research*, *205*(4), 577–581. https://doi.org/10.1007/s00221-010-2387-z

Nakano, T., Kuriyama, C., Himichi, T., & Nomura, M. (2015). Association of a nicotinic receptor gene polymorphism with spontaneous eyeblink rates. *Scientific Reports*, *5*(1), 8658. https://doi.org/10.1038/srep08658

Nakano, T., & Miyazaki, Y. (2019). Blink synchronization is an indicator of interest while viewing videos. *International Journal of Psychophysiology*, *135*, 1–11. https://doi.org/10.1016/j.ijpsycho.2018.10.012

Nakano, T., Yamamoto, Y., Kitajo, K., Takahashi, T., & Kitazawa, S. (2009). Synchronization of spontaneous eyeblinks while viewing video stories. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3635–3644. https://doi.org/10.1098/rspb.2009.0828

O'Sullivan, M. (2005). Emotional intelligence and deception detection: Why most people can't "read" others, but a few can. In R. E. Riggio & R. S. Feldman (Eds.), *Applications of Nonverbal Communication*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Raichle, M. E. (2015). The Brain's Default Mode Network. Annual Review of Neuroscience, 38, 433–447. https://doi.org/10.1146/annurev-neuro-071013-014030

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., Uchiyama, H., Kosaka, H., Okazawa, H., Fujibayashi, Y., & Sadato, N. (2010). “Stay tuned”: Inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience*, *4*, 127. https://doi.org/10.3389/fnint.2010.00127

Saito, K. (2024). Eye blink detection using personalized eye aspect ratio based on 3D landmark. In *Proceedings of the 38th Annual Conference of the Japanese Society for Artificial Intelligence* (4I3GS702). https://doi.org/10.11517/pjsai.JSAI2024.0_4I3GS702

Saito, K., Enomoto, M., Zenio, H., Hoshino, T., & Takahashi, E. (2025). Relationship between comprehension and blink synchronization. The 47th Annual Meeting of the Cognitive Science Society (CogSci2025).

Sato, W., Hyniewska, S., Minemoto, K., & Yoshikawa, S. (2019). Facial expressions of basic emotions in Japanese laypeople. *Frontiers in Psychology*, *10*, 259. https://doi.org/10.3389/fpsyg.2019.00259

Song, H., Park, B., Park, H., & Shim, W. M. (2021). Cognitive and Neural State Dynamics of Narrative Comprehension. Journal of Neuroscience, 41(43), 8972–8990. https://doi.org/10.1523/JNEUROSCI.0037-21.2021

Vrij, A., Hartwig, M., & Granhag, P. A. (2019). Reading lies: Nonverbal communication and deception. *Annual Review of Psychology*, *70*, 295–317. https://doi.org/10.1146/annurev-psych-010418-103135