Comparing Computer Vision Instruments for Eye Blink Analysis

Margaux Susman, Carla Miquel Blasco, Jan Bulla
University of Bergen
Bergen, Norway
margaux.susman@uib.no, carla.blasco@uib.no, jan.bulla@uib.no

Abstract

We compared four tools for analysing blink velocity and amplitude, examining how MediaPipe, OpenFace, InsightFace, and 3DDFA compare in terms of blink analysis. Building on previous findings that different tools yield different results (Kuznetsova and Kimmelman, 2024), we explored their fixed-effect estimates between linguistic versus non-linguistic blinks, within non-linguistic blinks (eye-watering blinks versus gaze-direction-change blinks), and within linguistic blinks (prosodic/turn-taking blinks, sign-aligned/list-marking blinks, and backchanneling blinks), while controlling for head pose (Pitch, Roll, Yaw). Using linear mixed-effects models on annotated French Sign Language data, we found tool-specific patterns: consistent negative effects for InsightFace and MediaPipe, but positive effects for 3DDFA. In addition, the influence of head pose varied across models (Pitch is strongly positive in MediaPipe but negative in InsightFace and some 3DDFA models; Roll and Yaw also vary in importance across tools). These discrepancies highlight methodological biases that can distort linguistic interpretations.

Keywords: blinks, computer vision, statistical analysis

1. Introduction

Computer vision tools are increasingly used in research on gestures and sign languages. In this context, they offer the possibility to track manual and non-manual movements, such as eye blinks, in a non-invasive and scalable way. Blinks constitute a particularly relevant object of study, as their timing and kinematic properties have been shown to relate to linguistic structure (Herrmann, 2010). Despite the growing use of computer vision techniques, little work has systematically evaluated how different tools compare when applied to fine-grained kinematic measures of non-manuals (Sargano et al., 2024; Kuznetsova and Kimmelman, 2024). In particular, potential tool-specific biases in the estimation of blink velocity and amplitude remain underexplored. The aim of this paper is to compare several computer vision tools: MediaPipe, OpenFace, InsightFace, and 3DDFA for the analysis of blink velocity and amplitude, using the same video data from the Dicta-Sign-LSF-v2 corpus (Belissen et al., 2020). The four tools examined in this study were selected to reflect established and more recent developments in computer vision. OpenFace and MediaPipe are widely used open-source frameworks for facial landmark detection and have been frequently adopted in linguistic and multimodal research, making them important reference points for comparison. InsightFace and 3DDFA, by contrast, represent more recent approaches that incorporate advances in deep-learning face analysis and 3D face alignment. We focus on blink velocity and amplitude across linguistically annotated blink functions. We contrast linguistic versus non-linguistic blinks, differences within non-linguistic blinks (eye-watering blinks versus gaze-direction-change blinks), and differences within linguistic blinks (prosodic/turn-taking, sign-aligned/list-marking, and backchanneling blinks) while controlling for head movements. Our evaluation reveals variation in measurements and statistical outcomes across tools. These include, e.g., higher estimated velocities for MediaPipe compared to OpenFace and reversed effect directions for 3DDFA in some models. We show that different tools exhibit different patterns and therefore highlight how measurement choices can shape the outcomes of blink kinematics analysis. We note that linguistic effects may partly reflect properties of the analysis pipeline rather than underlying behaviour.

2. Background

2.1. Blinks in Linguistic Research

Blinks are small-amplitude, high-frequency rapid closures (whether partial or complete) of the eyes. They are frequent in face-to-face interaction and have been studied in fields such as cognitive science, clinical research, and linguistics (Nyström et al., 2024; Herrmann, 2010). In research on gestures and sign languages in particular, blinks are of interest as non-manual events that may co-occur with discourse structure (e.g., marking turns and prosodic boundaries). In sign language research, it has been shown that blinks co-occur with lexical units as well. Because of their prominent role, blinks have been incorporated into sign language analyses since the mid-1990s. (Wilbur, 1994; Sze, 2004; Braffort and Chételat-Pelé, 2012; Herrmann,

Proceedings of the 12th Workshop on the Representation and Processing of Sign Languages, pages 459–467
16 May 2026.

2010).

At the same time, blinks offer a methodological challenge for empirical work. They are short, involve small movements, and are susceptible to interference from head movements and gaze direction. These properties make blinks particularly vulnerable to annotation uncertainty and measurement error, whether blinks are annotated manually or extracted automatically. As corpus-based and experimental linguistic studies increasingly rely on computer vision tools to scale up the analysis of non-manual behaviours, it becomes necessary to understand how methodological choices affect the resulting blink measures.

Despite the growing use of automated facial analysis in linguistic research, blink measurements have not yet been examined in a conclusive manner. Consequently, different analytical pipelines may yield different estimates of blink velocity and amplitude, with direct consequences for statistical modelling and linguistic interpretation. Without explicit comparison, it remains unclear whether reported effects reflect properties of the linguistic data or echoe noise and artifacts of the measurement tool.

2.2. Computer Vision Tools

This study compares four widely used computer vision frameworks for facial analysis: MediaPipe, OpenFace, InsightFace, and 3DDFA. Although all four tools enable the extraction of facial landmarks and head pose information, they rely on different modelling assumptions, training data, and tracking strategies, which may affect their performance on subtle and fast facial movements such as eye blinks.

MediaPipe is an open-source framework developed by Google that provides real-time, cross-platform pipelines for perception tasks (Lugaresi et al., 2019). Its Face Mesh model estimates 468 three-dimensional facial landmarks using a lightweight neural network optimized for speed and robustness. MediaPipe relies on a combination of face detection and landmark regression models trained on large-scale datasets, enabling stable tracking under varying lighting and pose conditions. Due to its dense landmark representation around the eyes, MediaPipe has become popular for fine-grained facial motion analysis, including blink detection (e.g., Delgado Gómez et al., 2024; Rahman et al., 2025).

OpenFace is an open-source toolkit designed for facial behaviour analysis in research contexts (Amos et al., 2016). It provides facial landmark detection, head pose estimation, eye gaze tracking, and facial action unit recognition. OpenFace is based on a constrained local neural field (CLNF) landmark detector, which uses statistical models combining local visual cues and general facial shape information to track facial landmarks consistently over time. Compared to MediaPipe, OpenFace uses a smaller set of facial landmarks but integrates them within a well-established facial behaviour analysis framework. Its emphasis on interpretability and reproducibility has led to widespread adoption in linguistics and social signal processing research (Zadeh et al., 2018).

InsightFace is a deep-learning-based framework primarily developed for face recognition and analysis (Guo et al., 2021). It includes modules for face detection, alignment, and landmark localisation based on state-of-the-art convolutional neural networks. InsightFace relies on large-scale training data and modern backbone architectures, resulting in high accuracy in face alignment tasks. While it is less commonly used in linguistic research, its strong performance in landmark detection renders it a promising candidate for blink analysis.

3DDFA (3D Dense Face Alignment) is a model-based approach that fits a three-dimensional morphable face model to two-dimensional images (Guo et al., 2020). It estimates dense 3D facial geometry and head pose by optimizing the parameters of a statistical face model. Unlike purely landmark-based systems, 3DDFA explicitly models facial shape and pose in three dimensions, which may improve robustness to large head movements. However, because the model is fitted separately to each frame, the measurements can vary slightly over time and be influenced by image noise. This is especially problematic for small and rapid movements such as eyelid motion.

These tools differ in their facial representations, ranging from dense landmark meshes (MediaPipe) and sparse landmark models (OpenFace) to deep alignment networks (InsightFace) and parametric 3D face models (3DDFA). They also vary in their optimisation objectives and how they smooth measurements over time. As a result, they may respond differently to rapid, low-amplitude movements such as blinks. Understanding these methodological differences is essential for interpreting between-tool variation in blink kinematic measures and for assessing the robustness of computer-vision-based analyses of non-manual behaviour.

3. Methods

We used the Dicta-Sign-LSF-v2 dataset (Belissen et al., 2020), which contains recordings of discussions on European travel. We annotated the blinks from 8 videos using the ELAN (version 6.2). We annotated a total of 1078 blinks and considered their functions. Only 650 blinks were analysed further, because they had a single function, whereas the remaining blinks had two or

more concurrent functions.

For each frame in each video, we then extracted the eye coordinates and head pose (which are known to influence facial coordinates, see Kuznetsova and Kimmelman, 2024) using each tool under comparison (MediaPipe, OpenFace, InsightFace, and 3DDFA) The 3D coordinates for each video were saved to CSV files. This step was implemented in Python. After that, we calculated the Eye Aspect Ratio (EAR) for each blink using the extracted coordinates. This ratio was introduced by Soukupova and Cech (2016) and is a measure of the degree of eye openness. It was computed from a fixed set of six eyelid landmarks per eye; for MediaPipe, we selected the closest available Face Mesh landmarks corresponding to the standard dlib eye contour (MediaPipe landmarks: left eye indices: 249, 362, 373, 380, 385, 387 and right eye indices: 7, 133, 144, 153, 158, 160). Then we applied the same EAR formula as for the other tools.

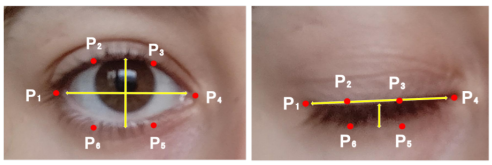

Figure 1: Eye landmark positions for the EAR calculation.

We merged the CSV files and removed rows without a blink and rows with multiple-function-blinks. For each blink, we retained only the first head-rotation measurement because of the short blink duration.

Subsequently, we derived blink kinematics (i.e. velocity and amplitude) from the smoothed EAR rather than from raw frame-by-frame EAR values in R (R Core Team, 2026; Bulla and Kimmelman, 2026). We selected smoothing splines as the smoothing method. More precisely, we relied on the function `ss` from the R package `npreg` (Helwig, 2020, 2024) to fit quintic smoothing splines via penalised least squares. The smoothing parameter λ was chosen by generalised cross-validation (GCV). For further details on the implementation, see (Bulla and Kimmelman, 2026). The smoothing served to reduce measurement noise and for capturing the overall eyelid movement during a blink. Blink amplitude reflects how far the eyelids closed at the blink peak and is expressed in units of EAR. Average blink velocity reflects how quickly the eyelids moved during the blink, and it is expressed as the average rate of change of EAR over time.

After computing these values, we added an annotation column based on the blink name, containing the annotation label. The labels were divided into linguistic and non-linguistic functions, and further into subfunctions. Linguistic subfunctions include: i) backchanneling blinks (feedback to an interlocutor), ii) prosodic blinks (boundary marking), iii) turn-taking blinks (beginning and end of a turn), iv) list-marking blinks, and v) blinks occurring with specific lexical signs. Prosodic and turn-taking blinks were grouped into one larger category, because both mark boundaries. In addition, sign-marking blinks and list-marking blinks were grouped together, because of their co-occurrence with specific signs. Non-linguistic blinks included eye-watering blinks and blinks induced by a change in gaze direction. We created three different datasets: for i) all blinks, ii) only large linguistic blink categories, and iii) only non-linguistic blinks. After this preparation, we carried out the statistical analysis of velocity and amplitude in R. There were six different analyses per tool tested: more specifically, we examined differences in amplitude and velocity for each of the three above-described datasets. The modelling procedure followed this logic: we fitted a baseline linear mixed-effects model (no fixed effects, random intercept for signer, estimated via maximum likelihood) using the `lme()` function from the `nlme` package (Pinheiro et al., 2025). Then we performed stepwise model selection by extending the baseline model towards more complex models, adding one predictor at a time. At each step, we selected the preferred model using the likelihood ratio test (LRT) for nested models and the AIC for non-nested comparisons. To improve residual normality, we applied a log or square-root transformation when necessary. As the linguistic models involved multiple comparisons, we used the package `marginaleffects` (Arel-Bundock et al., 2024) to perform post-hoc pairwise comparisons with Holm correction, thereby controlling the family-wise error rate. The processed blink kinematic measures and analysis scripts are available at the Open Science Framework (OSF): https://osf.io/r9yfs.

4. Results

In the following, we present the results for blink kinematics across our three linguistic categories for the four computer vision tools. We first report amplitude, then velocity.

4.1. Amplitude

Linguistic versus non-linguistic blinks - Table 1 presents estimated fixed effects from our selected models for predicting blink amplitude across the

four facial analysis tools considered ($N = 631$). The outcome scales vary: amplitude values were square-root transformed for the first three tools, and we applied a log transformation for 3DDFA.

We found that non-linguistic blinks show consistently reduced amplitude compared to linguistic ones: OpenFace ($p = 0.002$), MediaPipe ($p = 0.044$), 3DDFA ($p < 0.001$). Model selection did not include blink category for InsightFace. Head movement effects diverge across tools: Pitch is strongly positive in MediaPipe ($p < 0.001$), but negative for InsightFace ($p = 0.008$) and 3DDFA ($p < 0.003$). Yaw and Roll are positive in OpenFace (both $p < 0.001$), and Pitch, Yaw, and Roll are negative in 3DDFA ($p = 0.003, 0.031, 0.006$).

Non-linguistic blinks - Table 2 summarises fixed effects for the selected models predicting non-linguistic blink amplitude, i.e., differences between eye-watering blinks and gaze-direction-change blinks (baseline) across our four facial analysis tools ($N = 120$). We applied a square-root transformation for the OpenFace, MediaPipe and InsightFace data, while we used a logarithmic transformation for 3DDFA.

Where included, the blink category shows strong negative effects indicating reduced amplitude: InsightFace ($p < 0.001$) and OpenFace ($p < 0.001$). MediaPipe did not include the blink category (Yaw-only best model). Head movement effects diverge by tool: Pitch is positive in InsightFace ($p = 0.033$) but negative in OpenFace ($p = 0.011$); Yaw is positive in MediaPipe ($p < 0.001$); in 3DDFA, Roll ($p = 0.041$) and Yaw ($p = 0.004$) are negative.

Linguistic blinks - Table 3 summarises fixed effects for the best models predicting linguistic blink amplitude, i.e., the differences between backchanneling blinks (baseline), prosodic/turn-taking blinks, and sign-aligned/list-marking blinks across our four facial analysis tools ($N = 511$ observations).

Linguistic categories show tool-specific divergence. MediaPipe reveals clear reductions (prosodic/turn-taking, $p = 0.029$; sign-aligned/list-marking, $p = 0.027$). InsightFace trends show smaller reductions (prosodic/turn-taking, $p = 0.043$, sign-aligned/list-marking, $p = 0.069$). OpenFace shows no detectable effects ($p > 0.55$). 3DDFA reverses the pattern with positive coefficients (prosodic/turn-taking, $p < 0.001$; sign-aligned/list-marking, $p < 0.001$).

In Table 4, MediaPipe confirms its statistically significant results: backchanneling (3.72) > prosodic/turn-taking (2.65, Holm $p = 0.005$) > sign-aligned/list-marking (2.23, Holm $p = 0.034$). Other tools show non-significant trends post-correction.

As we have seen, non-linguistic blinks are generally smaller than linguistic ones for most tools, but the size and direction of this effect depend on the tool. InsightFace and MediaPipe tend to show reduced amplitude for prosodic/turn-taking and sign-aligned/list-marking blinks relative to backchanneling, OpenFace shows little categorical contrast, and 3DDFA often reverses the pattern. Head pose (Pitch, Roll, Yaw) also affects amplitude, but its sign and importance vary strongly by tool.

4.2. Velocity

The results for velocity analysis across our four tools are presented in this section.

Linguistic versus non-linguistic blinks - Table 5 shows fixed effects from our best models predicting blink velocity across our four tools ($N = 631$). The outcome scales vary: untransformed velocity (InsightFace), square-root-transformed velocity (OpenFace and MediaPipe) and log-transformed velocity (3DDFA).

We found that non-linguistic blinks show consistently reduced velocity compared to linguistic blinks where included: OpenFace ($p = 0.014$), MediaPipe ($p = 0.043$), and 3DDFA ($p < 0.001$).

Non-linguistic blinks - Table 6 summarises fixed effects from our best linear mixed-effects models, predicting non-linguistic blink velocity across our four tools. The outcome scales were a square-root transformation for InsightFace, OpenFace, and MediaPipe, and log transformation for 3DDFA.

Non-linguistic blink velocity analysis exhibits consistent patterns across tools. InsightFace and OpenFace both included the blink category in their best model and show strong negative effects for eye-watering blinks ($p < 0.001$), indicating that eye-watering blinks are reliably slower than gaze-direction-change blinks. MediaPipe and 3DDFA highlight head movement effects with Roll significant in both tools ($p < 0.001$ and $p = 0.020$, respectively).

Linguistic blinks - Table 7 summarises fixed effects for the best models predicting linguistic blinks velocity, specifically the differences between backchanneling blinks (baseline), prosodic/turn-taking blinks, and sign-aligned/list-marking blinks across the four tools ($N = 511$). In Table 7, we see a consistent pattern of slowing across the blink categories compared to backchanneling blinks, our baseline, but the strength and the direction of the effect depend on both the tool the and outcome scale used. The outcome scale varied depending on the tool: no transformation was applied for InsightFace, a square-root transformation was applied for OpenFace and MediaPipe, and a log transformation for 3DDFA. Across the three different tools for which we modelled blink categories (InsightFace, OpenFace, and MediaPipe), prosodic/turn-taking and sign-aligned/list-marking blinks tend to have

Predictor	InsightFace	OpenFace	MediaPipe	3DDFA
	(P + cat)	(Y + R + cat)	(P + cat)	(cat + P + R + Y)
Fixed effects				
(Intercept)	$0.267, p < 0.001$	$0.175, p < 0.001$	$0.497, p = 0.024$	$-0.491, p = 0.013$
Pitch	$-0.001, p = 0.008$		$17.657, p < 0.001$	$-2477, p = 0.003$
Yaw		$0.045, p < 0.001$		$-1315, p = 0.031$
Roll		$0.052, p < 0.001$		$-954, p = 0.006$
blink cat. (NOTLING)	$0.016, p = 0.016$	$-0.014, p = 0.002$	$-0.171, p = 0.044$	$-0.405, p < 0.001$

Table 1: Statistical results of the four tools for the comparison of linguistic versus non-linguistic blink amplitude. (P = Pitch, R = Roll, Y = Yaw, cat = blink category). **Colour coding:** Green = significant positive (p<0.05), Red = significant negative (p<0.05), Gray = non-significant.

Predictor	InsightFace	OpenFace	MediaPipe	3DDFA
	(cat + P)	(cat + P)	(R)	(R + Y + cat)
Fixed effects				
(Intercept)	$0.358, p < 0.001$	$0.215, p < 0.001$	$0.469, p < 0.001$	$-1.176, p < 0.001$
Pitch	$0.001, p = 0.033$	$-0.056, p = 0.011$		
Roll				$-3587, p = 0.041$
Yaw			$206.290, p < 0.001$	$-3099, p = 0.004$
blink cat. (WAT)	$-0.088, p < 0.001$	$-0.047, p < 0.001$		$-0.300, p = 0.111$

Table 2: Statistical results of the four tools for the comparison of non-linguistic blink categories (gaze-direction-change blinks (reference) and eye-watering blinks) on amplitude. (P = Pitch, R = Roll, Y = Yaw, cat = blink category). **Colour coding:** Green = significant positive (p<0.05), Red = significant negative (p<0.05), Gray = non-significant.

a lower velocity than backchanneling blinks (negative coefficients for sign-aligned/list-marking and prosodic/turn-taking blinks for InsightFace ($p = 0.003$ and $p = 0.050$ respectively), OpenFace (sign-aligned/list-marking $p = 0.016$) and MediaPipe (prosodic/turn-taking $p = 0.012$)). The InsightFace model suggests that backchanneling has the highest velocity, followed by prosodic/turn-taking blinks which are slightly slower and sign-aligned/list-marking blinks which are the slowest. Only the difference between backchanneling and sign-aligned/list-marking was statistically significant. OpenFace shows the same pattern as InsightFace but less pronounced: again, the difference between backchanneling and sign-aligned/list-marking is the only significant one. This suggests that OpenFace detects a similar slowing, but with weaker contrast. The MediaPipe model shows a clear separation between backchanneling and both prosodic/turn-taking and sign-aligned/list-marking categories: similar to InsightFace and OpenFace, backchanneling blinks have the highest velocity. The coefficients for prosodic/turn-taking (and to a lesser extent sign-aligned/list-marking) are negative and statistically significant. 3DDFA exhibits a completely different pattern. That is, prosodic/turn-taking and sign-aligned/list-marking show positive coefficients relative to backchanneling, implying higher velocity than baseline.

Overall, we note that non-linguistic blinks tend to have a slower velocity than linguistic ones (OpenFace/MediaPipe/3DDFA $p < 0.05$). Linguistic blinks show a backchanneling > prosodic/turn-taking > sign-aligned/list-marking gradient in InsightFace/OpenFace/MediaPipe, while we observe the opposite pattern with 3DDFA. For non-linguistic blinks, we note that eye-watering blinks are significantly slower than gaze-direction-change blinks (InsightFace/OpenFace $p < 0.001$).

5. Discussion

Our analyses show that the estimates of blink kinematics differ systematically across computer vision tools. In many cases, MediaPipe, OpenFace, and InsightFace agreed on effect direction, while 3DDFA tended to show weaker or opposite patterns, especially for linguistic blink categories. This shows that the linguistic effects on blink velocity and amplitude are not completely tool-independent, even when using identical video data and annotations.

The first implication is that we cannot assume that blink kinematic measures will behave identically across different analysis pipelines. In our data, non-linguistic blinks showed smaller amplitude and slower velocity than linguistic blinks in several tools. Linguistic categories showed the gradient: backchanneling > prosodic/turn-taking > sign-

Predictor	InsightFace	OpenFace	MediaPipe	3DDFA
	(P + cat)	(null)	(P + R + cat)	(cat + P + R)
Fixed effects				
(Intercept)	$0.270, p < 0.001$	$0.172, p = 0$	$1.072, p = 0.256$	$-0.748, p = 0.002$
Pitch	$-0.001, p = 0.006$		$56.383, p < 0.001$	$-2259, p = 0.008$
Roll			$355.472, p = 0.007$	$-459, p = 0.044$
Yaw				
blink cat. (PROS-TURN)	$-0.012, p = 0.043$		$-1.072, p = 0.029$	$0.599, p < 0.001$
blink cat. (SIGN-LIC)	$-0.014, p = 0.069$		$-1.496, p = 0.027$	$0.641, p < 0.001$

Table 3: Statistical results of the four tools for the comparison of linguistic blink categories (backchanneling (reference), prosodic/turn-taking & sign-aligned/list-marking blinks) on amplitude. (P = Pitch, R = Roll, Y = Yaw, cat = blink category). **Colour coding:** Green = significant positive (p<0.05), Red = significant negative (p<0.05), Gray = non-significant.

Tool	ADFE Mean (SE)	PROS-TURN Mean (SE)	SIGN-LIC Mean (SE)
InsightFace	$0.280\ (0.007)$	$0.269\ (0.007)$	$0.266\ (0.009)$
OpenFace	$0.172\ (0.015)$	$0.172\ (0.015)$	$0.176\ (0.015)$
MediaPipe	$3.720\ (0.927)$	$2.650\ (0.951)^*$	$2.230\ (1.052)$
3DDFA	$-0.888\ (0.235)$	$-0.289\ (0.238)$	$-0.247\ (0.252)$

Table 4: Post-hoc Pairwise Comparison (Response Scale, Holm Corrected) ($^*p < .01$ after correction; all others $p > .05$ except ADFE baselines), ADFE = backchanneling, PROS-TURN = prosodic/turn-taking, SIGN-LIC = sign-aligned/list-marking

aligned/list-marking blinks. However, the strength - and sometimes the direction - of the observed patterns varied by tool and, in some cases, by transformation (untransformed, square-root, log).

Second, the head-pose effects, which vary and are sometimes contradictory, show that controlling for pose is necessary but not sufficient: different tools estimate head pose in different ways. For example, some tools start from 2D landmarks, whereas 3DDFA fits a dense 3D morphable face model. Pitch contributed positively to amplitude and velocity in some MediaPipe models but negatively in others, and the relative importance of Roll and Yaw differed sharply between MediaPipe and 3DDFA. This suggests that head-pose predictors may compensate for tool-specific tracking behaviour rather than simply capturing signer movement.

The observed divergence across tools can also be attributed to their differences in modelling assumptions and facial representations. MediaPipe relies on a dense neural-network-based landmark mesh optimised for real-time robustness, OpenFace combines statistical shape constraints with landmark-based tracking, InsightFace employs deep alignment networks trained on large-scale datasets, and 3DDFA fits a parametric three-dimensional face model to each frame. Although we do not detail the exact causes, differences in tool design seem to affect blink estimates and statistical outcomes.

Note that our aim is to inform, not to endorse any specific tool as superior. We observed tool-dependent effects, so our study does not warrant strong recommendations about which tool should be favoured for blink analysis in sign language research. Instead, we encourage explicit reporting of the full analysis pipeline (tool versions, transformations, and model specifications).

Finally, this work has limitations that future research could address. Our sample is restricted to one sign language (LSF), a limited number of signers and recordings, and a particular set of kinematic measures derived from EAR-based tracking. Other annotation schemes and blink definitions might yield different results across tools. Extending the comparison to additional corpora could help determine the extent to which the tool dependence we observed generalises beyond the dataset considered here.

6. Author Contributions

The annotation and computer-vision processing and analysis were carried out by Margaux Susman, with input from Carla Miquel Blasco. The statistical analysis was performed by Margaux Susman under the supervision of Jan Bulla. The background on computer vision was drafted by Carla Miquel

Predictor	InsightFace	OpenFace	MediaPipe	3DDFA
	(P)	(P + cat)	(P + R + cat)	(cat + P + R + Y)
Fixed effects				
(Intercept)	$0.020, p < 0.001$	$0.094, p < 0.001$	$0.254, p = 0.008$	$-1.812, p < 0.001$
Pitch	$-0.000, p = 0.041$	$-0.011, p = 0.070$	$6.457, p < 0.001$	$-2809, p = 0.001$
Yaw				$-1048, p = 0.105$
Roll			$31.639, p < 0.001$	$-1036, p = 0.005$
blink cat. (NOTLING)		$-0.006, p = 0.014$	$-0.078, p = 0.043$	$-0.411, p < 0.001$

Table 5: Statistical results of the four tools for the comparison of linguistic versus non-linguistic blink velocity. (P = Pitch, R = Roll, Y = Yaw, cat = blink category). **Colour coding:** Green = significant positive (p<0.05), Red = significant negative (p<0.05), Grey = non-significant.

Predictor	InsightFace	OpenFace	MediaPipe	3DDFA
	(cat + P + R + Y)	(cat + P)	(R + P)	(R)
Fixed effects				
(Intercept)	$0.171, p < 0.001$	$0.105, p < 0.001$	$0.232, p < 0.001$	$-2.509, p < 0.001$
Pitch	$0.001, p = 0.003$	$-0.024, p = 0.027$	$2.818, p = 0.016$	
Roll	$0.000, p = 0.746$		$70.629, p < 0.001$	$-1208, p = 0.020$
Yaw	$-0.000, p = 0.996$			
blink cat. (WAT)	$-0.024, p < 0.001$	$-0.013, p < 0.001$		

Table 6: Statistical results of the four tools for the comparison of non-linguistic blink categories (gaze direction change blinks (reference) and eye-watering blinks) on velocity. (P = Pitch, R = Roll, Y = Yaw, cat = blink category). **Colour coding:** Green = significant positive (p<0.05), Red = significant negative (p<0.05), Grey = non-significant.

Blasco, with the other sections written by Margaux Susman. All authors commented on and approved the final manuscript.

7. Acknowledgements

Funded by the European Union (ERC, NONMANUAL, project number 101039378). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

8. Bibliographical References

Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. 2016. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2):20.

Vincent Arel-Bundock, Noah Greifer, and Andrew Heiss. 2024. How to interpret statistical models using marginaleffects for R and Python. *Journal of Statistical Software*, 111(9):1–32.

Valentin Belissen, Annelies Braffort, and Michèle Gouiffès. 2020. Dicta-Sign-LSF-v2: Remake of a continuous French Sign Language dialogue corpus and a first baseline for automatic sign language processing. In *12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 6040–6048, Marseille, France. European Language Resources Association (ELRA).

Annelies Braffort and Emilie Chételat-Pelé. 2012. Analysis and description of blinking in french sign language for automatic generation. In *Gesture and Sign Language in Human-Computer Interaction and Embodied Communication: 9th International Gesture Workshop, GW 2011, Athens, Greece, May 25-27, 2011, Revised Selected Papers 9*, pages 173–182. Springer.

Jan Bulla and Vadim Kimmelman. 2026. Processing kinematics of nonmanual markers in r.

John E Delgado Gómez, Elena Muñoz España, Carlos F Rengifo Rodas, and Diego E Guzmán Villamarín. 2024. Proposal for a eye blink detection using mediapipe, eye aspect ratio and peak identification. In *International Conference on NeuroRehabilitation*, pages 345–349. Springer.

Jia Guo, Jiankang Deng, Alexandros Lattas, and Stefanos Zafeiriou. 2021. Sample and compu-

Predictor	InsightFace	OpenFace	MediaPipe	3DDFA
	(cat + P)	(cat)	(P + cat + R)	(cat + P + R)
Fixed effects				
(Intercept)	$0.021, p < 0.001$	$0.089, p < 0.001$	$0.282, p = 0.012$	$-2.067, p < 0.001$
Pitch	$-0.000, p = 0.065$		$7.154, p < 0.001$	$-2707, p = 0.003$
Roll			$22.166, p = 0.023$	$-687, p = 0.005$
blink cat. (PROS-TURN)	$-0.002, p = 0.050$	$-0.003, p = 0.088$	$-0.090, p = 0.012$	$0.595, p < 0.001$
blink cat. (SIGN-LIC)	$-0.003, p = 0.003$	$-0.006, p = 0.016$	$-0.091, p = 0.066$	$0.615, p < 0.001$

Table 7: Statistical results of the four tools for the comparison of linguistic blink categories (backchanneling (reference), prosodic/turn-taking & sign-aligned/list-marking blinks) on velocity. (P = Pitch, R = Roll, Y = Yaw, cat = blink category). **Colour coding:** Green = significant positive (p<0.05), Red = significant negative (p<0.05), Grey = non-significant.

Tool	ADFE Mean (SE)	PROS-TURN Mean (SE)	SIGN-LIC Mean (SE)
InsightFace	$0.0215\ (0.0013)$	$0.0199\ (0.0013)$	$0.0182\ (0.0015)$
OpenFace	$0.0895\ (0.0067)$	$0.0864\ (0.0067)$	$0.0836\ (0.0069)$
MediaPipe	$0.576\ (0.110)$	$0.485\ (0.112)$	$0.485\ (0.116)$
3DDFA	$-2.23\ (0.216)$	$-1.63\ (0.220)$	$-1.61\ (0.238)$

Table 8: Post-hoc pairwise comparison (response Scale, Holm Corrected, ($^*p < .01$ after correction; all others $p > .05$ except ADFE baselines), ADFE = backchanneling, PROS-TURN = prosodic/turn-taking, SIGN-LIC = sign-aligned/list-marking

tation redistribution for efficient face detection. *arXiv preprint arXiv:2105.04714*.

Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. 2020. Towards fast, accurate and stable 3d dense face alignment. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Nathaniel Helwig. 2020. Multiple and Generalized Nonparametric Regression. In *SAGE Research Methods Foundations*. SAGE.

Nathaniel E. Helwig. 2024. *npreg: Nonparametric Regression via Smoothing Splines*.

Annika Herrmann. 2010. The interaction of eye blinks and other prosodic cues in german sign language. *Sign Language & Linguistics*, 13(1):3–39.

Kuznetsova and Kimmelman. 2024. Testing mediapipe holistic for linguistic analysis of nonmanual markers in sign languages. *arXiv preprint arXiv:2403.10367*.

Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172*.

Marcus Nyström, Richard Andersson, Diederick C Niehorster, Roy S Hessels, and Ignace TC Hooge. 2024. What is a blink? classifying and characterizing blinks in eye openness signals. *Behavior Research Methods*, 56(4):3280–3299.

José Pinheiro, Douglas Bates, and R Core Team. 2025. *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-168.

R Core Team. 2026. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Anima Rahman, Roger Woodman, and Valentina Donzella. 2025. Advancing blink detection in driver monitoring with improved eye landmark analysis.

Allah Bux Sargano, Sébastien Vandenitte, Tommi Jantunen, and Vadim Kimmelman. 2024. Evaluation of head pose estimation algorithms for sign language analysis. In *2024 International Conference on IT and Industrial Technologies (ICIT)*, pages 1–6. IEEE.

Felix Sze. 2004. Blinks and intonational phrasing in hong kong sign language. In *Signs of the time: Selected papers from TISLR*, pages 83–107.

Ronnie Wilbur. 1994. Eyeblinks & ASL phrase structure. *Sign Language Studies*, 84:221–240.

Amir Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. 2018. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2236–2246, Melbourne, Australia. Association for Computational Linguistics.