Investigating calibration constraints for the processing of a narrow-view multi-camera system

Petra Helmholz[1], Derek Litchi[2], Liam Boyle[1], Richard Palmer[3], Roslyn Ward[3]

[1] School of Earth and Planetary Sciences, Curtin University, Kent St, Bentley, Australia (Petra.Helmholz; Liam.Boyle)@curtin.edu.au
[2] Department of Geomatics Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (ddlichti@ucalgary.ca)
[3] Curtin School of Allied Health, Curtin University, Kent St, Bentley, Australia (R.Ward; R.Palmer)@curtin.edu.au

Vision Metrology (WG II/9)

Keywords: Camera Calibration, Least Squares Adjustment, Calibration Constraints.

Abstract

Speech is a highly complex and multidimensional process, requiring precise coordination of muscular actions within the vocal tract. Disruptions or delays in speech motor control often lead to speech impairments. Recent advancements in markerless facial tracking technology enable the collection of objective measurements to assess these impairments. To obtain such photogrammetric measurements, a multi-camera network is employed, making accurate camera calibration essential. This paper examines the constraints applied during the calibration process. Two adjustment strategies were evaluated. The first, Independent Adjustment (IDP), performs self-calibration for each camera without introducing constraints. The second, Combined Adjustment (CMB), incorporates object space constraints by ensuring that object point locations observed from all cameras remain consistent. Given the cameras’ narrow fields of view, both IDP and CMB were tested with additional constraints related to the principal point offset. Each adjustment was executed under two conditions: fixing the principal point offset to zero or estimating it as part of the calibration. Results indicate that the choice of adjustment significantly affects the interior orientation parameters (IOPs). IDP with the principal point offset fixed to zero produced the most accurate outcomes. However, variations in IOPs had no meaningful impact on object space coordinates. These findings suggest that the simplest approach—IDP with the principal point offset fixed to zero—offers reliable calibration for multi-camera systems used in speech assessment. This streamlined method can be adopted in future applications to enhance efficiency without compromising accuracy.

1. Introduction

Speech is a complex, multidimensional act, produced through precisely timed and controlled coordination of muscular actions within the vocal tract (Smith and Zelaznik, 2004). When speech motor control is disturbed or delayed, these disruptions may result in compensatory articulatory movements and contribute to speech impairments (Kent, 2000). Researchers have observed such compensatory behaviours in children with speech impairments, including jaw instability, lip asymmetry and undifferentiated tongue movements (Orton et al., 2025; Mogren et al., 2022; Kabakoff et al., 2023).

Accurate assessment of speech impairments requires evaluating the relative contribution of cognitive, linguistic and speech motor control (Strand and McCauley, 2008). Despite this need, clinicians currently lack access to suitable tools for such assessments. Until now, our understanding of compensatory speech movements has been derived from instrumentation techniques that are largely limited to research laboratories (Menard and Tiede, 2024). However, advancements in markerless facial tracing technology now allow for the collection of objective measurements outside of laboratory settings, thus expanding the potential for clinical application (Palmer et al., 2024).

To this end, in 2024, Palmer et al. reported on the utilisation of computer vision-based techniques to extract clinically relevant measurements of articulatory movements. Specifically, their application demonstrated the feasibility of extracting jaw and lip measurements in speech from 2D video recordings. Further work also demonstrated that the objective kinematic measures extracted also show good agreement with perceptual judgements (Orton, et al., 2025).

This present study extends the work of Palmer et al. (2024) by applying markerless facial tracking and computer vision-based measurement within a multi-camera network. By addressing multiple viewpoints, the study addresses the calibration and constraint challenges inherent to single-camera systems, aiming to enhance both the precision and clinical applicability of computer vision techniques for the assessment of speech motor control.

To generate objective and automatic scores, the output from the multi-camera network used to capture the videos is required to produce good and measurable quality results. Past investigations focused on:

- The theoretical design of the multi-camera network (Boyle et al., 2024)
- Variance propagation of the image processing to the clinically relevant measurements of articulatory movements (Boyle et al., 2025),

The camera network consisted of three stationary time-synchronised cameras labelled BMC, BMR and BML (Figure 1). Boyle et al. (2024) characterised the best design as a maximum horizontal angle of ±45° for BMR and BML with BMC directly facing the participant. Furthermore, the vertical angle changes of +15° of the horizontal plane defined by the head position and for BML and BMR at a vertical angle of -15°. To ensure that both side-view cameras (BMR/BML) could see the entire front of the face, the maximum horizontal angle was set to approximately ±35°. For practical reasons, the cameras were all set to horizontal plane defined by the head position of the participant. Boyle et al. (2025) introduced variance propagation to derive objective measurements relevant to the assessment of speech impairment.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-657-2026 | © Author(s) 2026. CC BY 4.0 License.

For the variance propagation introduced in Boyle et al., (2025), each camera was calibrated individually following a standard procedure.

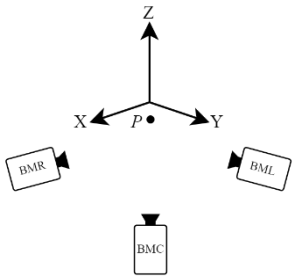

Figure 1: Visualisation of the three camera arrangements comprising BMR (capturing the right and front of the face), BMC (capturing the front of the face), and BML (capturing the left and front of the face) in 3D space converging on a singular point (*P*). (Boyle et al., 2024)

Boyle et al.,(2024, 2025) did not include an investigation of the different constraints that can be applied to the least-squares adjustment of the multi-camera network. In this publication, we address the gap and investigate the constraints that can be applied to the camera calibration process, particularly given the long focal lengths of the cameras. Four different constraints are investigated and validated using several methods, including the precision of the camera calibration adjustments and the impact of the derived camera calibration parameters on observations in object space.

The paper is structured as follows: The background is presented in section 2. The methodology is introduced in section 3, followed by the dataset in section 4. Results are presented in section 5. The manuscript closes with a discussion and a conclusion in section 6.

2. Background

Constraining a bundle adjustment is not novel and has been reported in several publications. Youcai and Haralick (1999) constrained the configuration of control points in object space, e.g., forcing points to be on a line, to have line-to-plane angle constraints or to constrain distances between the points. Constraints to the control point locations were also performed by Huang et al. (2024), where control points were constrained to lie on a planar surface. Pacheco and Tommaselli (2024) minimised the resulting parallax error in omnidirectional images to establish a value for the projection cylinder radius as close as possible to the object's depth.

Alternatively, constraints to the exterior orientation parameters (EOPs) of multi-camera systems can be introduced. Lee and Yilmaz (2010) constrained the EOPs (X^C,$Y^C Z^C$, ω, φ and κ) of an aerial multi-head camera system without platform calibration parameters and navigation solutions by adding them as additional observations in the bundle adjustment. Huang et al. (2022) and Lerma et al. (2010) constrained the base of cameras in a multi-camera system to be fixed. Detchev et al. (2017) investigated four estimation methods of constraining camera mounting parameters (CMPs). The methods ranged from independent estimation of the Interior Orientation Parameters (IOPs) and EOPs, followed by constraining the CMPs, to in-situ calibration of all unknowns, simultaneously adding geometric constraints to the mathematical model. The in-situ calibration of all parameters yielded the most practical and strongest adjustment solutions, driven by the increased number of observations, the reduced number of unknowns, and the improved network geometry. Similar findings were concluded by Helmholz and Lichti (2020), who investigated three constraints on the EOPs. The independent (IDP) method imposes no constraints. The combined (CMB) method constrains the locations of common object points captured by the camera array. Finally, the EOPs are constrained in addition to the CMB constraint. In this paper, we follow the procedure as introduced by Helmholz and Lichti (2020).

When performing network orientation with self-calibration on images with very narrow fields of view, Stamatopoulos and Fraser (2011) reported problems arising from over-parameterisation, ill-conditioning, and subsequent numerical instability in the normal equations of the bundle adjustment. They demonstrated that much of the instability in the solution of the traditional self-calibrating bundle adjustment for narrow-angle photography can be attributed to an incomplete formulation of the partial derivatives of the extended collinearity equations with respect to the principal point parameters. The proposed expansion of the terms for x_p and y_p in the configuration matrix **A** will mitigate the solution instability to a considerable extent. A complete parameterisation of x_p and y_p is applied in our work.

Kenefick et al. (1972) also reported challenges when processing very narrow-view cameras. Their proposed solution was to reduce the parameters of the interior orientation, such as setting zero higher-order radial lens distortion parameters, as well as the principal point offset x_p and y_p. In this paper, we investigate the constraint of setting the principal point offset x_p and y_p zero.

The complexity of processing images with long focal lengths in close-range photogrammetry led us to the decision not to use off-the-shelf photogrammetric solutions or computer vision implementations. To find the best constraints is an essential requirement to move to the next step of the project that will focus on applying congruency testing of facial inter-landmark distance as a measure of speech production and motor control. Hence, we aim to carefully investigate the network adjustment constraints using a more traditional approach.

3. Method

3.1 Camera specification and set-up

Blackmagic Pocket Cinema 4K cameras are capable of resolutions up to 4096×2160 and 60 frames per second and are therefore suitable for analysing speech movements. Three such cameras were set up approximately 3.0-3.5m from the participants. Each was combined with an Olympus Digital 45mm (f1.8) prime lens. The camera/lens combination defined the FOV and was selected to meet the lab's space constraints while allowing for close-ups of the face. As in Boyle et al. (2024), the primary camera (BMC) was placed front on. The other cameras were positioned on either side of BMC to capture the face's left and right sides (BML and BMR, respectively), as shown in Figure 1. The three BM cameras were time-synchronised based on the detection of an audible signal. The synchronisation accuracy is within one frame and was confirmed by inspection of discrete movements visible across all cameras.

3.2 Functional Model

Photogrammetric processing of images assumes that all light rays travel in straight lines and upon entering the lens system, form a bundle that passes through the perspective centre. This means an object point i $(X,Y,Z)_i$, its homologous image point $(x,y)_{ij}$ in

https://doi.org/10.5194/isprs-annals-XI-2-2026-657-2026

image j, and the perspective centre of the image j $(X,Y,Z)_j^C$ lie on a straight line in a three-dimensional space. The collinearity equations are given by

$$x_{ij} + \hat{v}_{x_{ij}} = x_{p_j} - c_j \frac{U_{ij}}{W_{ij}} + \Delta x_{ij} \quad (1)$$

$$y_{ij} + \hat{v}_{y_{ij}} = y_{p_j} - c_j \frac{V_{ij}}{W_{ij}} + \Delta y_{ij} \quad (2)$$

where $\hat{v}_x$, $\hat{v}_y$ are the image point residuals, x_p, y_p are the coordinates of the principal point offset, c is the principal distance, and Δx, Δy are the additive correction terms as any deviation from collinearity must be modelled by introducing additional camera parameters. These parameters are typically referred to as the IOPs. In this paper, we consider the radial lens distortion parameters Δx_R,Δy_R only. Furthermore, (U_{ij},V_{ij},W_{ij}) are formulated as

$$\begin{pmatrix} U_{ij} \\ V_{ij} \\ W_{ij} \end{pmatrix} = \mathbf{M} \left(\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_i - \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}_j^C \right) \quad (3)$$

where $\mathbf{M}$ is the matrix used for the rotation from object space to image space that can be parameterised as

$$\mathbf{M} = \mathbf{R}_3(\kappa_j)\mathbf{R}_2(\varphi_j)\mathbf{R}_1(\omega_j) \quad (4)$$

and ω, φ and κ are the rotation angles around the X, Y and Z axes, respectively. (X^C,Y^CZ^C) together with ω, φ and κ per image/camera station are often referred to as the EOPs for each camera station.

3.3 Least-Squares Adjustments and Constraints

The parametric (or Gauss–Markov) adjustment model was used for the self-calibration (Brown, 1971). All IOPs are modelled as network invariants. Datum definition was with inner constraints on object points. All unknowns (EOPs, IOPs and object points) were estimated simultaneously. Constraints added to the solution were implemented as weighted constraints according to the unified approach to least squares (Mikhail, 1976). Two different types of least-squares solution adjustments were implemented as described in Table 1.

Independent	**IDP**	No added constraints. Cameras are self-calibrated independently.
Combined	**CMB**	The locations of common object points are constrained to match.

Table 1: Overview of the used adjustment types (left column), the abbreviation used in this paper (centre column) and a brief explanation (right column).

Both adjustments were performed in two modes: solving for the principal point offset (x_p and y_p) (solved) and setting it to zero (fixed), resulting in the adjustment modes shown in Table 2.

	Adjustment mode	Adjustment type	Principal point offset (x_p and y_p)
1.	IDP_PPO_fixed	IDP	Set to 0,0
2.	IDP_PPO_solved	IDP	Solved
3.	CMB_PPO_fixed	CMB	Set to 0,0
4.	CMB_PPO_solved	CMB	Solved

Table 2: Calibration tests performed.

3.4 Validation

3.4.1 Comparison of IOPs based on Adjustments

The IOPs based on the calibration tests presented in Table 2, were assessed. The differences in the principal distance, principal point offset and radial lens distortions were also assessed.

3.4.2 Impact of IOPs on object space observations

For our application, it is important to analyse the impact of the different derived IOPs on the 3D object-space coordinates. Two methods were used to quantify the impact of IOPs in object space: a calibration field, and a participant's facial landmarks. For both methods, the derived IOPs were first used to calculate the room-based EOPs of each camera.

Resection: The resection process calculates the EOP (X^C,Y^CZ^C, ω, φ and κ) per camera station j from the 2D image (x_i, y_i) of a room-centred calibration wall. The wall control object points i (X,Y,Z) were calibrated in the standard manner. Using the calculated IOPs of each camera j, the collinearity equations (equations 1 and 2) were used to estimate the EOPs using the parametric (or Gauss–Markov) adjustment model. The linearised collinearity equations for all EOPs are grouped into the standard parametric model

$$\mathbf{v} = \mathbf{A} \cdot \Delta\mathbf{x} - \mathbf{l}, \quad (7)$$

with $\mathbf{A}$ being the design matrix of partial derivatives taken with respect to the EOPs, $\mathbf{l}$ being the vector of image point observations, $\Delta\mathbf{x}$ being the vector of corrections for the unknowns $\mathbf{x}$ (EOPs) and $\mathbf{v}$ being the residuals. $\mathbf{A}$ is formulated as

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_{x_{ij}}}{\partial X^C{}_j} & \frac{\partial f_{x_{ij}}}{\partial Y^C{}_j} & \frac{\partial f_{x_{ij}}}{\partial Z^C{}_j} & \frac{\partial f_{x_{ij}}}{\partial \omega_j} & \frac{\partial f_{x_{ij}}}{\partial \varphi_j} & \frac{\partial f_{x_{ij}}}{\partial \kappa_j} \\ \frac{\partial f_{y_{ij}}}{\partial X^C{}_j} & \frac{\partial f_{y_{ij}}}{\partial Y^C{}_j} & \frac{\partial f_{y_{ij}}}{\partial Z^C{}_j} & \frac{\partial f_{y_{ij}}}{\partial \omega_j} & \frac{\partial f_{y_{ij}}}{\partial \varphi_j} & \frac{\partial f_{y_{ij}}}{\partial \kappa_j} \end{bmatrix}. \quad (8)$$

Intersection: The method described by Boyle et al. (2025) was adopted to verify the impact of different IOPs on observations in object space, first using a calibration field and then using a participant's facial landmarks. As in Boyle et al. (2025), the BlazeFace algorithm was used to automatically extract facial landmarks.

The intersection process calculates the 3D coordinates (X_i, Y_i, Z_i) for each point/landmark i from the 2D image (x_i, y_i) observations. This was done by applying the collinearity equations, utilising the previously calculated IOPs and EOPs of each camera j (equations 1 and 2).

The parametric (or Gauss–Markov) adjustment model was also used for the intersection. The linearised collinearity equations for all image points in a block can be grouped into the standard parametric model (equation 7) with $\mathbf{A}$ being the design matrix of partial derivatives taken with respect to the object space point coordinates, $\mathbf{l}$ being the vector of image point observations, $\Delta\mathbf{x}$ being the vector of corrections for the unknowns $\mathbf{x}$ (3D object space coordinates) and $\mathbf{v}$ being the residuals. $\mathbf{A}$ is formulated as

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_{x_{ij}}}{\partial X_i} & \frac{\partial f_{x_{ij}}}{\partial Y_i} & \frac{\partial f_{x_{ij}}}{\partial Z_i} \\ \frac{\partial f_{y_{ij}}}{\partial X_i} & \frac{\partial f_{y_{ij}}}{\partial Y_i} & \frac{\partial f_{y_{ij}}}{\partial Z_i} \end{bmatrix}. \quad (9)$$

The associated weight matrix for this group of observations was denoted as $\mathbf{P}$. The image weights were set to 3 pixels based on the findings in Boyle et al. (2025).

4. Dataset and processing

The overall workflow is described in Figure 2. The camera IOPs were extracted using the different constraints introduced in section 3.3. The cameras' EOPs were determined using resection, and the image observations of the facial markers were extracted using a markerless algorithm. These datasets were then used in an intersection to find the 3D object of the facial markers which can be validated comparing to a reference captured with the VectraH2 camera. All steps are introduced in detail below.

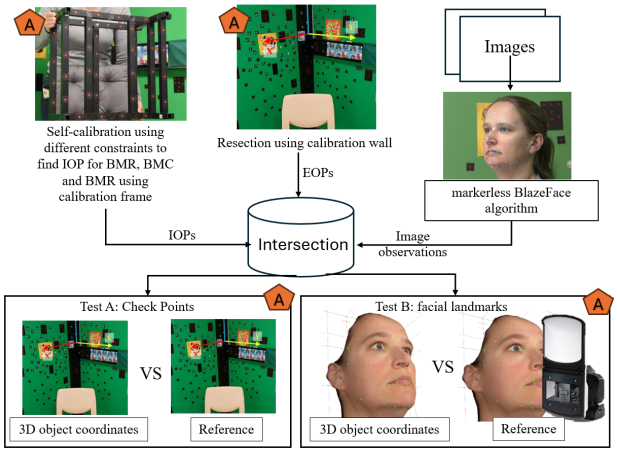

Figure 2: Data processing workflow. Validations/assessments are performed for the steps highlighted with the orange A symbol.

4.1 Interior Orientation Parameters (IOPs) determination using IDP and CMB self-calibration using calibration cube

The long focal lengths of the BM cameras required them to remain stationary during calibration, which made it impossible to use the room-based calibration system. Instead, the calibration frame shown in Figure 2 was moved in front of the cameras, approximately at the location marked *P* in Figure 1. The frame incorporated 95 ground control points (GCPs) in three planes. Prior to the capture, the location on the GCPs were determined using an in-situ calibration utilising a Nikon D750 SLR Nikon with a Nikkor 24-70mm f/2.8 ED G AF-S lens and a calibrated scale bar. The achieved Root Mean Square (RMS; coordinate precision) of this calibration is provided in Table 3; with the overall RMS being 0.15 mm. The precision was in the sub-mm range and sufficient for our assessment. The frame was rotated and tilted, providing good geometry for solving for the IOPs.

	X-axis	Y-axis	Z-axis
RMS [mm]	0.09	0.08	0.09

Table 3: Summary of the RMS coordinate precision from the bundle adjustment for calibration cube.

4.2 Exterior Orientation Parameters (EOPs) determination using resection of calibration wall GCPs

A total of 292 GCPs defined the room-centred coordinated system (Figure 2). The GCP coordinates were determined using the same D750 SLR Nikon with a Nikkor 24-70mm f/2.8 ED G AF-S lens and scale bar. The origin of the coordinate system of the GCPs was placed approximately at head height, as shown in Figure 2 (top centre). The achieved RMS for the calibration of the control field is shown in Table 4 with the overall RMS being 0.15 mm. The precision in the sub-mm range is sufficient for assessment. The average EOPs and their standard deviations are shown in Table 10. BMR mainly contributed to the large standard deviations in the X and Y axes due to poor object space configuration, and hence must be considered when analysing the results.

	X-axis	Y-axis	Z-axis
RMS [mm]	0.09	0.08	0.09

Table 4: Summary of the RMS coordinate components from the bundle adjustment for all wall and scale points.

4.3 Image observations of facial landmarks

As in Boyle et al. (2025), we combined the markerless BlazeFace algorithm with photogrammetry. BlazeFace was used to infer the positions of 45 clinically relevant facial landmarks, mapped to anthropometric points. Figure 3 shows the location of the landmarks on the face. For validation, the facial landmarks of a child were used as they are the focus of the study (due to ethical constraints, one of the authors’ faces is used for demonstrative purposes herein). For each camera station, these landmarks were extracted from a short video sequence of a child participant (523) with a neutral facial expression. For analysis, we used three frames extracted from the short video sequence; one from the start, one from the middle, and one from the end. Photogrammetric intersection was used to derive the 3D coordinates of these facial landmarks.

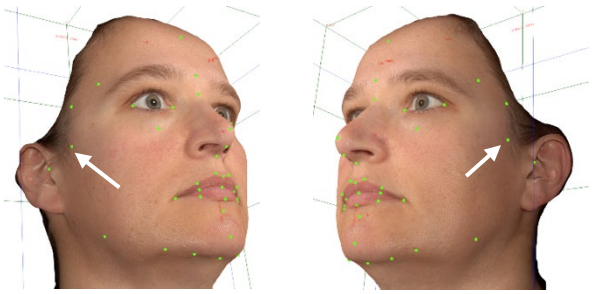

Figure 3: Placement of the clinically relevant landmarks. Some landmarks that are difficult to place by visual inspection are highlighted with a white arrow.

4.4 Validation

An intersection was performed using the IOPs, EOPs, and image observations of Check Points from the calibration wall (Test A) and of facial landmarks (Test B).

For Test B, the VectraH2 camera (Canfield Scientific, NJ, USA) was used to capture the participant’s face. Immediately after capturing the video, the VectraH2 image was taken, obviating the possibility of facial change. The 3D image was imported into Cliniface (Palmer et al., 2020), and a trained researcher placed the landmarks as shown in Figure 3. These points formed the reference dataset used to validate accuracy. A 3D similarity transformation (fixed scale) was performed on the intersected points relative to the VectraH2 reference landmarks to account for the different coordinate systems used. The results of the similarity transformations were used to evaluate accuracy.

5. Results

The different adjustments were assessed to determine their success. Subsequently, the IOPs calculated using the different constraints were compared and the impact of the different IOPs on object space observations were evaluated. Finally, the extracted facial landmark locations based on the different IOPs

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-657-2026 | © Author(s) 2026. CC BY 4.0 License.

were compared to the matched reference image captured using the VectraH2 camera.

5.1 Interior Orientation Parameters (IOPs) determination using IDP and CMB self-calibration

5.1.1 *Adjustment results*: Adjustments were successful, as shown in Table 5 and Table 6. The RMS values of the image point residuals (Table 5) are all in sub-mm range. The largest value is 0.012 mm (highlighted in red). RMS values of the object space coordinates are in sub-mm range (Table 6) with the largest value being 0.52 mm (also highlighted in red).

			RMS x	RMS y
IDP	PPO fixed	BML	0.010	0.010
		BMC	0.011	0.010
		BMR	0.010	0.008
	PPO solved	BML	0.010	0.010
		BMC	0.011	0.010
		BMR	0.010	0.010
CMB	PPO fixed		0.005	0.005
	PPO solved		**0.012**	0.009

Table 5: Image point residuals after the LSA. All RMS values in [mm].

			RMSx	RMSy	RMSz
IDP	PPO fixed	BML	0.70	0.56	0.52
		BMC	0.18	0.19	0.13
		BMR	0.29	0.28	0.30
	PPO solved	BML	0.70	0.56	**0.52**
		BMC	0.18	0.19	0.28
		BMR	0.52	0.13	0.30
CMB	PPO fixed		0.08	0.11	0.07
	PPO solved		0.09	0.10	0.10

Table 6: Mean Object Space Coordinates after the LSA. All RMS values in [mm].

5.1.2 *Principal Distance*: The principal distances of the different adjustments are shown in Table 7. The large standard deviations between 0.25 mm and 0.50 mm (highlighted in red) can be explained by the narrow FoV, which causes correlations with the EOPs. BMC had the shortest principal distance across all adjustments (with an average of 120.944 mm), and BML had the longest (with an average of 122.159 mm). The range for BML was 1.011 mm, for BMC 0.874 mm, and for BMR 0.785 mm. The differences between the different adjustment resolutions are not meaningful for practical purposes.

			c [mm]	σc [mm]
IDP	PPO fixed	BML	122.531	0.389
		BMC	120.663	0.314
		BMR	121.899	0.498
	PPO solved	BML	122.406	0.397
		BMC	120.752	0.315
		BMR	121.682	**0.499**
CMB	PPO fixed	BML	121.520	0.347
		BMC	121.537	0.280
		BMR	121.609	0.384
	PPO solved	BML	122.181	0.272
		BMC	120.825	**0.249**
		BMR	121.114	0.315

Table 7: Comparison of the principal distances and their standard deviations based on the adjustments.

5.1.3 *Principal point offset*: The principal point offsets of the different adjustments are shown in Table 8. Again, the large standard deviations (between 0.198 mm and 0.403 mm (highlighted in red) are explainable by the narrow FoV, which causes correlations with the EOPs. The principal point offsets are all very large and significantly different from zero. Across the adjustments, BMR had the largest displacement (-2.016 mm), and BMC the smallest (-0.092 mm); both highlighted in blue. The radial lens distortion parameters were not significant and are therefore not presented.

			xP [mm]	σXp [mm]	yP [mm]	σYp [mm]
IDP	PPO solved	BML	1.125	0.319	0.574	**0.403**
		BMC	0.132	0.259	1.282	0.326
		BMR	-1.603	0.334	0.951	0.381
CMB	PPO solved	BML	1.660	0.215	0.339	0.293
		BMC	**-0.092**	**0.198**	1.360	0.267
		BMR	**-2.016**	0.264	0.486	0.275

Table 8: Comparison of the principal point offset based on the adjustments.

5.2 Exterior Orientation Parameters (EOPs) determination using resection

The camera locations were resected relative to a room coordinate system. The resection was performed using Australis (Fraser and Edmundson, 2000). The same well-distributed 13 points were used for the resection of each camera. The resection results (image coordinate x and y residuals) are shown in Table 9, and the EOPs with their standard deviations in Table **10**: Average EOPs and average σ for the EOPs of all camera stations. Locations are in [mm], angles in [°].Table 10. All image coordinate residuals in Table 9 show sub-pixel accuracy, ranging from 0.46 pixels to 0.84 pixels (highlighted in red). Table 10 show the average values and standard deviations of the EOPs after the resection. The values for BML and BMC are in the acceptable range, but BMR show weak results (highlighted in red). Investigations indicate a poor object-point configuration. Their impact will be considered for the analysis of the results.

			#pts	Image res x	Image res y	Image res total
IDP	PPO fixed	BML	13	0.56	0.64	0.60
		BMC	13	0.58	0.66	0.62
		BMR	13	**0.46**	0.56	0.51
	PPO solved	BML	13	0.56	0.70	0.64
		BMC	13	0.82	0.65	0.74
		BMR	13	0.48	0.57	0.53
CMB	PPO fixed	BML	13	0.59	0.65	0.62
		BMC	13	0.53	0.68	0.61
		BMR	13	0.56	0.80	0.69
	PPO solved	BML	13	0.62	0.71	0.66
		BMC	13	0.54	0.69	0.62
		BMR	13	0.61	**0.84**	0.74

Table 9: Results of calibration. All RMS values in [pixels].

	X	**Y**	**Z**	**ω**	**φ**	**κ**
	σ_X	σ_Y	σ_Z	σ_ω	σ_φ	σ_κ
BML	1286.7	3130.9	159.0	-90.1	19.2	-178.9
	4.9	0.3	2.9	0.4	0.6	0.1
BMC	2908.4	2386.7	149.5	-91.8	51.3	-175.7
	4.0	0.5	5.9	0.8	0.5	0.1
BMR	3313.8	517.8	-0.4	-125.8	83.8	-140.7
	17.6	11.8	7.7	3.6	5.5	0.6

Table 10: Average EOPs and average $\boldsymbol{\sigma}$ for the EOPs of all camera stations. Locations are in [mm], angles in [°].

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-2-2026-657-2026 | © Author(s) 2026. CC BY 4.0 License.

5.3 Test A: 3D object coordinates assessment using Check Points

Five points visible in the field of view of all three cameras were intersected, and the resulting 3D coordinates were compared with the reference dataset. The resulting RMS values are shown in Table 11. The best result was achieved by IDP with the PPO fixed (0.15 mm, highlighted in red), and the worst result was 0.22 mm for CMB with the PPO fixed (also highlighted in red).

	IDP		CMB	
RMS	PPO fixed	PPO solved	PPO fixed	PPO solved
x	0.08	0.07	0.12	0.09
y	0.04	0.03	0.09	0.05
z	0.14	0.15	0.16	0.15
Total	**0.15**	0.16	**0.22**	0.18

Table 11: Intersection result compared to reference points. All RMS values in [mm].

5.4 Test B: 3D object coordinates assessment using Facial Landmarks

This test was performed in two parts:

- Test B1: Comparison against the best-performing network in the self-calibration
- Test B2: Comparison against the VectraH2 reference

5.4.1 *Test B1: Comparison against the best-performing network in the self-calibration*: The landmarks shown in Figure 3 were extracted from three frames within a video sequence. The frames are placed early in the production of a word (frame 1), in the middle (frame 15) and at the end (frame 30) of a word. The previously calculated IOPs and EOPs were used in an intersection to determine the 3D landmark locations for each adjustment type and frame. The previously assessed best-performing adjustment type (IDP, PPO fixed) was used as a reference to compare against the intersection results from all other adjustment types (Table 12). The results for frames within a given adjustment type are consistent, and there are hardly any differences. Overall, RMS values are in the sub-mm range and vary only minimally across adjustment types. The largest RMS total values were observed for the CMB PPO-solved adjustment (highlighted in red). Hence, the different adjustments show no practical impact in object space.

		Frame	RMS x	RMS y	RMS z	RMS Total
IDP	PPO solved	1	0.02	0.01	0.00	0.02
		15	0.02	0.01	0.00	0.02
		30	0.02	0.01	0.00	0.02
CMB	PPO solved	1	0.15	0.10	0.03	**0.18**
		15	0.15	0.10	0.03	**0.18**
		30	0.14	0.10	0.03	**0.18**
	PPO solved	1	0.02	0.01	0.00	0.02
		15	0.02	0.01	0.00	0.02
		30	0.02	0.01	0.00	0.02

Table 12: Intersection result compared to IDP PPO fixed. All RMS values in [mm].

5.4.2 *Test B2: VectraH2 comparison:* The RMS values of the 3D similarity transformations from each intersection dataset to the VectraH2 coordinate systems are shown in Table 13. All solutions have large RMS values, with the CMB PPO fixed having the best results by a margin (1.629 mm, highlighted in red).

	IDP		CMB	
RMS	PPO fixed	PPO solved	PPO fixed	PPO solved
x	0.062	0.057	0.064	0.057
y	1.292	1.293	1.292	1.293
z	0.998	0.999	0.991	0.999
Total	1.634	1.634	**1.629**	1.634

Table 13: Intersection result compared to reference points. All RMS values in [mm].

Figure 4 shows plots of the residuals in the x/y and the x/z plane evaluated on a single face. All adjustment modes have very similar differences to the VectraH2 – indeed so similar that the residual plot of *CMB, xp,yp + solved* completely masks the residuals of the other adjustment types. A trend is evident: landmarks on the outer face show the largest displacements. The residuals have clear, consistent, and laterally symmetric directions.

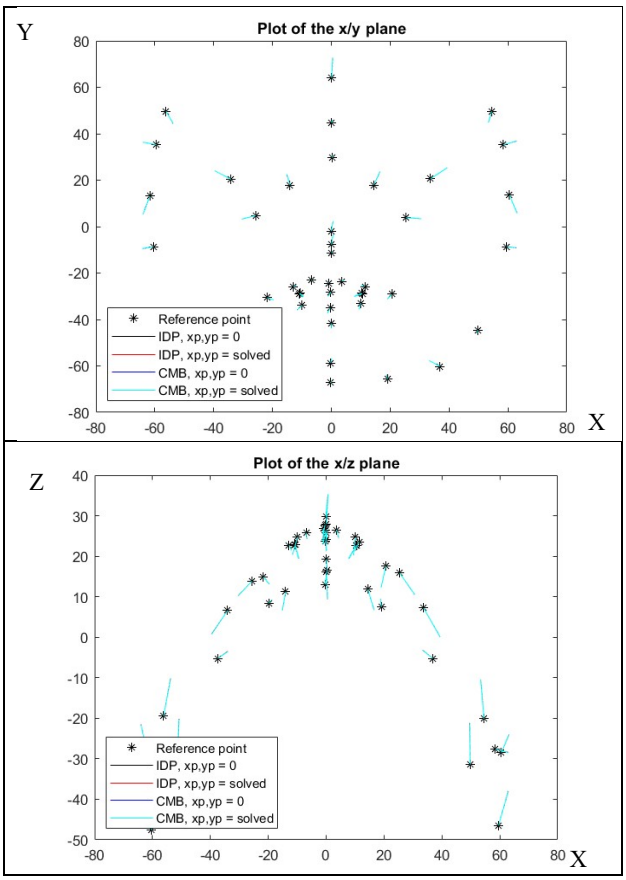

Figure 4: Residual plots of the intersection results vs VectraH2 reference.

5.5 Discussion

Several reasons could contribute to the observed large residuals:

1. Poor accuracy in the photogrammetric intersection causes depth and scaling errors of landmarks on the outer face
2. Erroneous landmark placement by the expert within the VectraH2 reference dataset
3. Incorrect correspondence of BlazeFace points to clinically relevant landmarks
4. Incorrect placement of facial landmarks by BlazeFace

Each of these are considered and discussed in turn.

5.5.1 *Poor accuracy in the photogrammetric intersection*: Results presented in Table 11 showed very small RMS values (sub-mm) for the Check Point assessment when using the calibration wall. If poor accuracy in the photogrammetric intersection had been present, the values in Table 11 would have been much larger. Hence, we conclude that the large residuals are not due to the photogrammetric workflow, but rather to one or a combination of the other factors.

5.5.2 *Erroneous landmark placement by the expert within the VectraH2 reference dataset*: Variability in landmark placement was observed within the VectraH2 reference dataset. Landmarks located on well-defined features, such as those around the mouth, showed smaller residuals (Figure 4), indicating these landmarks may be easier to place with greater precision. However, this was not consistent across all well-defined facial features as landmarks around the eyes showed large residuals. It was therefore concluded that the landmark placement variability by the expert within the VectraH2 reference dataset contributed to the residuals but not of the magnitude observed in Table **13**. However, it must be noted that some of the facial landmarks are difficult for a user to place in the VectraH2 dataset by visual inspection alone, without the ability to palpate the region, as indicated by the white arrows in Figure **3**.

5.5.3 *Incorrect correspondence of BlazeFace points to clinically relevant landmarks*: BlazeFace extracted a total of 468 facial landmarks of which 45 clinically relevant facial landmarks were selected for this work. It is possible that some clinically relevant landmarks were not tracked by a corresponding BlazeFace landmark. However, given the density of the BlazeFace mesh, total RMS values of 1.6 mm (Table **13**), we consider this unlikely. Though not within the scope of this study, evaluating how the locations of the BlazeFace facial points align with these landmarks warrants further investigation.

5.5.4 *Incorrect placement of facial landmarks by BlazeFace*: To investigate the cause of the large residuals, the extracted BlazeFace points used for the intersection were overlaid with an image extracted from the video sequence (Figure 5). The incorrect placement of the BlazeFace points on the facial perimeter is clearly visible in regions furthest away from the camera locations (red arrows in Figure 5) as well as some mid-face points (orange arrows in Figure 5). Initial investigation shows a strong correlation of displacement errors with increasing side view of the face. However, this analysis was conducted on a single face, and the nature of the observed residuals may differ across a larger sample of individuals and may not show a consistent pattern. While relative measurements, as used in Palmer et al. (2024), may not be affected, this will be a concern for absolute measurements. Further investigations are required.

6. Conclusion

This paper investigated constraints on the camera calibration process for the processing of 3D facial landmarks to support the assessment of speech sound disorders. Two different types of least-squares solution adjustments were implemented. The first adjustment type, referred to as the independent adjustment (IDP), involved the self-calibration of the cameras being performed separately. No constraints were added to this adjustment. The second adjustment included constraints on object space. The combined adjustment (CMB) constrained the locations of object points observed by each camera to be identical.

The previous literature on narrow-field-of-view cameras has suggested constraining the principal point to the centre of the image. Hence, the IDP and CMB calibrations were performed using two modes: a) adjustment for the principal point offset, and b) fixing its location to the centre of the image.

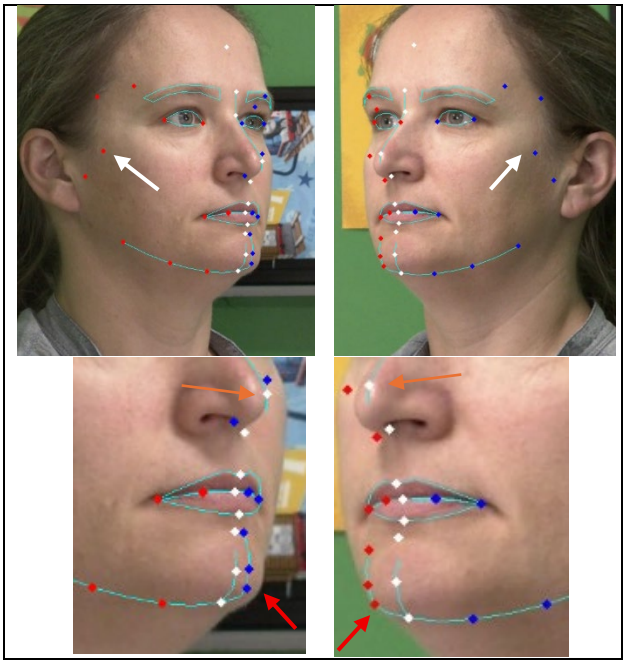

Figure 5: BlazeFace points overlayed with image of BML (on the right) and BMR (on the left). Some landmarks that are difficult to place by visual inspection are highlighted with a white arrow. Incorrectly placed points are highlighted with orange and red arrows.

The results show significantly different IOPs, with the IDP adjustment fixing the principal offset to zero, producing marginally better results. The different IOPs had no practical effect on the object-space coordinates. Future work applying deformation analysis for speech assessment should use the simplest adjustment: IDP fixing the principal point offset to zero.

More critical to accuracy is BlazeFace's extraction of facial image points, which are used for the intersection. Our initial investigations show large discrepancies in landmark placement for the side-viewing cameras (BMR/BML). While relative measurements of a front-on view camera, as used in Palmer et al. (2024), may not be affected, for photogrammetric applications further investigation is needed since intersection results are greatly affected. A more systematic analysis of the impact of viewing direction on the BlazeFace algorithm is needed and will be performed in future work.

Acknowledgements

Research reported in this publication was supported by the Medical Research Future Fund under grant number 2016518 2021 MRFF Chronic Neurological Conditions and PROMPT Institute Research Grant 59592.

References

Alsadik, B. S., Gerke, M., Vosselman, G., 2012. Optimal Camera Network Design for 3d Modeling of Cultural Heritage. ISPRS Ann. *Photogramm. Remote Sens. Spatial Inf. Sci.*, I-3, 1-7. doi.org/10.5194/isprsannals-I-3-7-2012

Boyle, L., Helmholz, P., Lichti, D. D., Ward, R., 2024. Validation of Camera Networks Used for the Assessment of Speech Movements, *ISPRS Archives., XLVIII-2-2024*, 41–48, https://doi.org/10.5194/isprs-archives-XLVIII-2-2024-41-2024, 2024.

Boyle, L., Helmholz, P., Ward, R., Palmer, R., Lichti, D., 2025. Uncertainty analysis of clinically relevant distances derived using photogrammetric intersection for the assessment of motor-speech-control in children, *ISPRS Archives, XLVIII-G*-2025, 225–232, https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-225-2025, 2025.

Brown, D.C.,1971. Close range camera calibration. *Photogram. Eng.* 1971, 37, 855–866.

Detchev, I., Habib, A., Mazaheri, M., Lichti, L., 2017. Practical In Situ Implementation of a Multicamera Multisystem Calibration. *Journal of Sensors*. Volume 2018, Article ID 5351863, 12 pages. https://doi.org/10.1155/2018/5351863

Fraser, C.S. and Edmundson, K., 2000. Design and Implementation of a Computational Processing System for Off-Line Digital Close-Range Photogrammetry. *ISPRS Journal of Photogrammetry & Remote Sensing*, 55(2): 94-104, 2000.

Helmholz, P. and Lichti, D.D., 2020. Investigation of chromatic aberration and its influence 2 on the processing of underwater imagery. *Remote Sens. 2020, 12, 3002; doi:10.3390/rs12183002*

Huang, D., Elhashash, M., Qin, R., 2022. Constrained Bundle Adjustment for Structure From Motion Using Uncalibrated Multi-Camera Systems. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume V-2-2022.

Huang, W., Miao, H., Jiao, S., Miao, W., Xiao, C., Wang, Y., 2024. A planar constraint optimization method to improve camera calibration for imperfect planar targets. *Optics and Lasers in Engineering* 180 (2024) 108273

Kabakoff, H., Beames, S. P., Tiede, M., Whalen, D. H., Preston, J. L., and McAllister, T., 2023. Comparing metrics for quantification of children’s tongue shape complexity using ultrasound imaging. *Clinical linguistics & phonetics*, 37(2), 169-195.

Kenefick, J.F., Gyer, M.S., Harp, B.F., 1972. Analytical Self-Calibration. *Photogrammetric Engineering*, pp. 1117 – 1126.

Kent, R. D., 2000. Research on speech motor control and its disorders: A review and prospective. *Journal of Communication disorders,* 33(5), 391-428.

Lee, Y-L., Yilmaz, A., 2010. Temporal Geometric Constrained Bundle Adjustment for the Aerial Multi-Head Camera System. *ISPRS Archives* – Volume XXXVIII Part 4, 2011

Lerma, J.L., Navarro, S., Cabrelles, M., Segui, A.E., 2010. Camera Calibration with Baseline Distance Constraints. *The Photogrammetric Record* 25(130): 140–158 (June 2010).

Lichti, D.D., Jarron, D., Tredoux, W., Shahbazi, M., 2020. Geometric Modelling and Calibration of a Spherical Camera Imaging System. *The Photogrammetric Record* 35(170): 123–142 (June 2020). DOI: 10.1111/phor.12315

Ménard, L. and Tiede, M., 2024. Instrumental Analysis of Speech Production. *The Handbook of Clinical Linguistics*, Second Edition, 489-504.

Mikhail, E.M.,1976. Observations and Least Squares; *IEP*, New York, USA: 1976; 497p. ISBN-13: 978-0819123978.

Mogren, Å., McAllister, A., Sjögreen, L., 2022. Range of motion (ROM) in the lips and jaw during vowels assessed with 3D motion analysis in Swedish children with typical speech development and children with speech sound disorders. *Logopedics Phoniatrics Vocology,* 47(4), 219-229.

Orton, L., Palmer, R., Ward, R., Helmholz, P., Strauss, G. R., Davey, P., Hennessey, N. W., 2025. Discriminating Children with Speech Sound Disorders from Children with Typically Developing Speech Using the Motor Speech Hierarchy Probe Words: A Preliminary Analysis of Mandibular Control. *Diagnostics*, 15(14), 1793.

Palmer, R. L., Helmholz, P., Baynam, G., 2020. Cliniface: Phenotypic Visualisation and Analysis using Non-Rigid Registration of 3D Facial Images, *ISPRS Archives*., XLIII-B2-2020, 301–308, https://doi.org/10.5194/isprs-archives-XLIII-B2-2020-301-2020, 2020

Pacheco, J.M. and Tommaselli, A.M.G., 2024. Simultaneous Calibration of Multiple Cameras and Generation of Omnidirectional Images. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume X-1-2024.

Palmer, R., Ward, R., Helmholz, P., Strauss, G. R., Davey, P., Hennessey, N., Namasivayam, A., 2024. Facial Movements Extracted from Video for the Kinematic Classification of Speech. *Sensors*, 24(22), 7235.

Smith, A. and Zelaznik, H. N., 2004. Development of functional synergies for speech motor coordination in childhood and adolescence. *Developmental psychobiology*, 45(1), 22-33.

Stamatopoulos, C, and Fraser, C., 2011. Calibration Of Long Focal Length Cameras in Close Range Photogrammetry. *The Photogrammetric Record* 26(135), pp 339–360 (September 2011). DOI: 10.1111/j.1477-9730.2011.00648.x

Strand, E. A. and McCauley, R. J., 2008. Differential diagnosis of severe speech impairment in young children. T*he ASHA Leader,* 13(10), 10-13.

Youcai, H. and Haralick, R.M., 1999: Testing camera calibration with constraints. *Photogrammetric Engineering and Remote Sensing* 65(3): 249-258