

Validation of Camera Networks Used for the Assessment of Speech Movements

Liam Boyle¹, Petra Helmholz¹, Derek D. Lichti², Roslyn Ward³

¹ School of Earth and Planetary Sciences, Curtin University, Kent St, Bentley, Australia (Liam.Boyle; Petra.Helmholz}@curtin.edu.au

² Department of Geomatics Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada (ddlichti@ucalgary.ca)

³ School of School of Allied Health, Curtin University, Kent St, Bentley, Australia (R.Ward@curtin.edu.au)

Commission II/WG1

Keywords: Photogrammetry, Network Design, Speech Motor Control.

Abstract

The term speech sound disorder describes a range of speech difficulties in children that affect speech intelligibility. Differential diagnosis is difficult and reliant on access to validated and reliable measures. Technological advances aim to provide clinical access to measurements that have been identified as beneficial in diagnosing speech disorders. To generate objective measurements and, consequently, automatic scores, the output from multi-camera networks is required to produce quality results. The quality of photogrammetric results is usually expressed in terms of the precision and reliability of the network. Precision is determined at the design stage as a function of the geometry of the network. In this manuscript, we focus on the design of a photogrammetric camera network using three cameras. We adopted a similar workflow as Alsadika et al. (2012) and tested several network configurations. As the distances from the camera stations to object points were fixed to 3500mm, only the horizontal and vertical placements of the cameras were varied. Horizontal angles were changed within an increment of 10°, and vertical angles were changed within an increment of 5°. The object space coordinates of GCPs for each camera configuration were assessed in terms of horizontal error ellipses and vertical precision. The best design was the maximum horizontal and vertical convergence angles of 90° and 30°. The existing camera network used to capture videos for speech assessment was approximately as good as the top third of tested designs. However, from a validation perspective, it can be concluded that the design is viable for continued use.

1. Introduction

Speech sound disorders (SSDs) are the most common childhood communication impairment, representing approximately 70% of a paediatric speech-language pathologist's (S-LP's) caseload (Dodd, 2014). SSDs can have lifelong impacts; therefore, timely access to intervention is critical to minimise negative outcomes (Daniel & McLeod, 2017).

Clinical assessment is pivotal in establishing a diagnosis and facilitating timely access to targeted intervention. Current clinical assessment practices for identifying and diagnosing SSDs typically involve perceptual analyses to identify speech error patterns and processes (McLeod and Baker, 2014). However, these methods of analysis are typically subjective. Whilst objective measures are available within the research setting, they are neither accessible nor feasible within the clinical setting.

With technological advances, the reporting on video-based technologies, developed to provide objective speech kinematic measurements, has increased (Jafari et al., 2023). These technologies aim to provide clinical access to measurements that have been identified as beneficial in diagnosing speech disorders but are typically constrained to the laboratory setting (Jafari et al., 2023). One such technology under development is the Speech Movement and Acoustic Analysis Tracking (SMAAT, www.smaat.org) tool that leverages computer vision and machine learning techniques to generate objective and automatic scores from multimodal analysis of video recordings captured from a network of multiple cameras.

To generate objective and automatic scores, the output from the multi-camera network used to capture the videos is required to produce quality results. As stated by Fraser (1984), the quality of photogrammetric results is usually expressed in terms of the precision and reliability of the network. Precision is determined at the design stage as a function of the geometry of the network. Reliability is concerned with controlling the quality of conformance of an observed network to its design. This manuscript will focus on the precision of the photogrammetric network design for the SMAAT project. We aim to investigate and validate the best camera network by designing various network configurations and determining the optimal configuration using photogrammetry techniques. These results will be compared to the existing network configuration used within SMAAT.

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

2. Background

Camera network designs are not unique and have been implemented in various scenarios and fields of research. Following the classification scheme of Grafarend (1985), the interrelated problems of network design can be identified as:

- Zero-Order Design: the datum problem.
- First-Order Design: the configuration problem,
- Second-Order Design: the weight problem, and
- Third-Order Design: the densification problem.

For this manuscript, we focus on the first-order design, that is, the design of the photogrammetric camera network configuration. Datum definition, in which the network position, orientation, and scale are defined in the local coordinate system, is defined by pre-surveyed control points. Thus, the datum definition is not of concern to the SMAAT project. The second-order design is currently addressed by an equally weighted simulation process. Further investigations into the second- and third-order design aspects will be covered in future work.

Saadat et al. (2004) grouped first-order design, the network configuration design, into three classes: accessibility-related constraints, visibility, and range.

Accessibility-related constraints are typically dependent upon physical constraints of space, obstructions, and often the infeasibility of occupying specific geometrically favourable locations. This constraint applies to all applications and is the reason why camera network designs are the focus of several publications (Alsadik et al., 2012). The network constraints for the SMAAT project will be introduced in the method section. For instance, the number of cameras is currently limited to three.

Visibility-related constraints come from the visibility of a cluster of object points from a camera station, which depends upon the constraints of target incidence angle, occluded areas, and the camera's field of view (FOV). This constraint is not frequently discussed in photogrammetric applications but more in surveillance. For instance, Erdem and Sclaroff (2006) investigated the camera placement problem to satisfy coverage constraints while minimising the total cost by applying a radial sweep algorithm. While an interesting investigation, the geometry of the camera locations was purely optimised for coverage but not to derive precise 3D measurements. Nevertheless, occlusions exist when capturing facial landmarks from different viewing directions. Hence, the camera network must ensure that at least two cameras cover each region of interest of the face. A constraint that is predicted to be met easily when using a multi-camera network.

Range or distance-related constraints include those applying to imaging scale, resolution, FOV, depth of field (DOF), number and distribution of points and workspace (object space). The FOV, DOF and resolution are fixed for the SMAAT application due to the utilised cameras and lens. Any simulation must use the camera specifications as defined for the SMAAT project. The distribution of points in object space is defined by the facial landmarks of interest. The points must be head-centred to account for movement.

Range constraints are typically divided into two parts: the minimum and maximum distance of the cameras to the object (Saadat et al., 2004). This range constraint was investigated by Aminia (2017), who focused on the range of the camera stations presented by the exterior orientation parameters (EOPs) and object coordinates, including the angle between the camera viewing direction and the object surface. Aminia's contribution was establishing a decision system using fuzzy computation that can identify unsuitable images based on network design constraints that may have an unfavourable effect on the result of bundle adjustment. This approach is post-capturing and processing and has several camera stations available for the analysis.

In addition to the range constraint identified by Saadat et al. (2004), consideration must also be given to the distance between the camera stations (Zhao et al., 2008). For instance, while

Ahmadabadian et al. (2014) were constrained by two cameras with a fixed very short baseline, the image pairs captured were not limited. Barazzetti (2017) also addressed baseline constraints, focusing on the network designs with short baselines between the images. Short baselines are a challenge for sequences when object space points are only covered in a small set of images. A larger number of object points does not significantly improve precision: the main constraint is the poor network of the camera stations. In contrast, closed solutions (images captured in a cylinder layout) produce results of magnitude improvements even if the same camera base and a similar number of points were used.

Mathematical solutions for optimal camera placement have been investigated, including the work of Fraser (1984) and Olague and Mohr (2002). Alsadika et al. (2012) and Buschinelli et al. (2020) both based their network design on Fraser (1984). The first step in these papers is to create a sparse point cloud model of the object of interest. The sparse point cloud allows an initial set of camera stations to be defined, followed by an optimisation by filtering the block to achieve a minimal quantity of camera stations required, and finally, an optimisation to confirm the design. Optimisation of the remaining camera EOPs is undertaken by a defined function to minimise the average coordinate errors of the point cloud in the initial phase. We will adopt a similar workflow.

Finally, a constraint often not included in publications related to camera network designs is that they must be easy to set up, calibrate, and use for capture by non-photogrammetric professionals. So, the network must be effective and relatively simplistic, and minor variation should not lead to large quality reductions.

3. Methodology

Each step is guided by the constraints defined within the SMAAT project. Hence, these constraints are introduced first.

Next, the steps listed below, as described by Alsadika et al. (2012), are used to find the optimal camera network:

- Defining object space points
- Defining camera locations
- Finding the optimal camera locations

3.1 SMAAT constraints

3.1.1 Accessibility-related constraints: Currently, SMAAT uses a fixed camera network of 3 cameras. They are placed in front of the participant (camera named BMC) and approximately 45° to each side (cameras are named BML for the camera capturing the participant's left side and BMR for the camera capturing the participant's right side). A visualisation of the camera arrangement is shown in Figure 1 relative to other components associated with this paper. All simulations will assume a network using these three cameras. The same labelling will be applied.

3.1.2 Range or distance-related constraints: The distance constraint is defined according to:

1. The FOV of the used cameras
2. The space available in the lab for data capture

The three Blackmagic Pocket Cinema cameras are set up approximately 3.0-3.5m from the participants. They are combined with Olympus Digital 45mm (f1.8) prime lenses. The

camera/lens combination defines the FOV, and it has been selected to fit the space constraint in the lab and simultaneously allow the capture of faces as close-ups. Blackmagic Pocket Cinema 4K cameras are capable of resolutions up to 4096×2160 and 60 frames per second and, therefore, suitable for analysing speech movements.

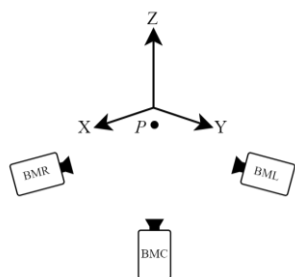


Figure 1: Visualisation of the three camera arrangements comprising BMR, BMC, and BML in 3D space converging on a singular point (P) representing a participant.

The cameras were calibrated using a customised calibration frame and following a standard photogrammetric workflow, ensuring a strong geometry. The Interior Orientation Parameters (IOPs) of the cameras are summarised in Table 1. The average focal length (121.5269 mm) was calculated and used for the simulations. Notably, the focal length c is relatively long compared to the image format size (very narrow FOV). Subsequently, the radial lens and decentring distortion parameters were found to be insignificant, so they are neither reported here nor used in any of the simulations. Note that while the interior model comprises the focal length and the principal point offsets, only the former are reported here due to their strong impact on the network design. For the purpose of simulation, the latter can be assumed to pass through the exact centre of the image plane.

	BML	BMC	BMR
c [mm]	121.6382	121.3302	121.6124
σc [mm]	0.3575	0.3088	0.3899

Table 1: Focal length parameter of the three Blackmagic cameras.

3.2 Object space points for the network design simulations

In contrast to the sparse point cloud methodology of Alsadika et al. (2012), we utilise a combination of red retroreflective and white coded circular Ground Control Points (GCPs), which are placed on the walls behind the chair used for the capture of participant's faces (Figure 2). A total of 292 GCPs were tested. The GCP coordinates were determined by calibrating the field using a Nikon D750 SLR Nikon with a Nikkor 24-70mm f/2.8 ED G AF-S lens. The origin of the coordinated system of the GCPs is placed approximately in head height, as shown in Figure 2. The scale was defined by a single scale constraint by a fixed length scale bar. The achieved RMS for the calibration of the control field is provided in Table 2; the overall RMS is 0.154 mm. The precision is in the sub-mm range and is sufficient for our network design assessment.

	X-axis	Y-axis	Z-axis
RMS [mm]	0.090	0.083	0.095

Table 2: Summary of the RMS coordinate components from the bundle adjustment for all wall and scale points.

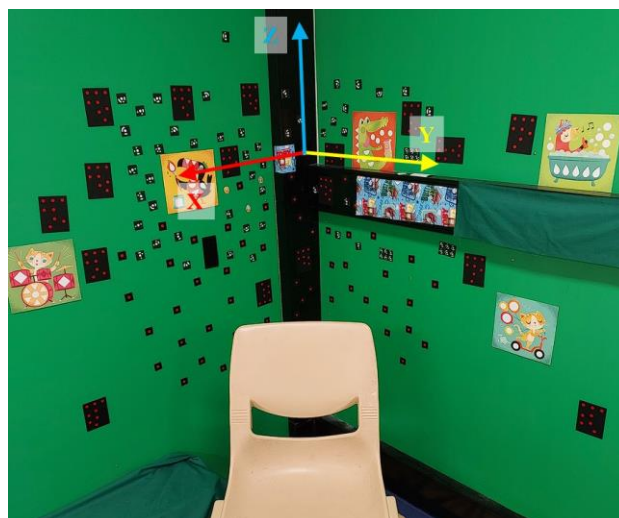


Figure 2: Approximate participant seating location with GCPs in the background.

To have a realistic representation of object points comparable to what is used when facial landmarks are captured, not all wall GCPs are used for the network design. Instead, a reduced number of control points approximately in the location of the head is used, as shown in Figure 3. Figure 3 shows 60 GCPs as simulated in image space; however, depending on the design, the number of GCPs used varies between 37 and 69.

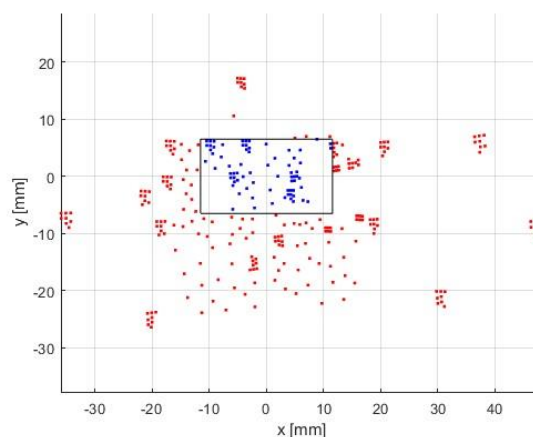


Figure 3: Example of the approximate selection of GCPs in image space used for the simulations as seen by a single central camera.

3.3 Defining camera locations

Simulated camera locations: As the distances from the camera stations to the object points were fixed to 3500mm, only the horizontal and vertical placements of the cameras could be varied. Horizontal angles (θ) were changed in increments of 10° , and vertical angles (ϵ) were changed within an increment of 5° (Figure 4, Table 3). The horizontal convergence angles varied between a minimum of 10° and a maximum of 90° . The vertical convergence angles were more constricted with a minimum and maximum of 0° and 30° .

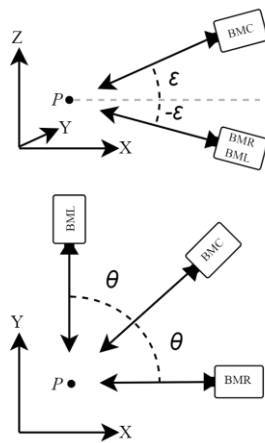


Figure 4: Diagrammatic representation of horizontal angle θ and elevation angle ϵ applied in determining EOP parameters.

As only three camera stations were available, each network design used a combination of only three camera stations. In all designs, the middle camera (BMC) is directly placed front onto the participant/reduced GCP field.

Furthermore, the lab space presents a physical limitation such that the maximum horizontal displacement of the cameras has been limited to 90° which theoretically prevents BMR and BML from contacting the surrounding walls. From a simulation perspective, this means the widest horizontal design is comparable to the existing SMAAT design. The similarity can be seen in Figure 5 (top) where the SMAAT design (pink) creates an approximate 80° convergence angle; a minor disadvantage in horizontal geometry compared to the maximum tested angle. As per Figure 5 (bottom), the SMAAT design also incorporates asymmetrical vertical height change. To reflect this but in a more controlled manner, the simulations place the cameras in various vertical positions at all horizontal locations.

We investigated several “flat designs” where all camera stations were positioned in one horizontal plane and at equal height to the location of the head. The left and right cameras were equidistant to the centre Blackmagic Camera (BMC). The maximum horizontal angle to the side cameras is $\pm 45^\circ$, hence forming a 90° angle between the camera from the far right to the far left. All “flat designs” are highlighted in Table 3. Vertical angles were less extensively tested, the total maximum vertical angle was capped at 30° . These designs are presented in Table 3. The decision to minimise the total vertical angle to any angle below 15° is based on the requirement for facial landmarks to be present in at least two images. Increasingly steeper vertical angles will reduce the number of facial landmarks that can be extracted due to occlusions.

All camera stations used for the simulation are shown in blue in Figure 5. In the same figure, the GCPs are shown in green and provide the context of the camera locations relative to the wall GCPs and the simulated convergence point is shown in black. The convergence point is intended to replicate the theoretical position of an adult participant's head during data capture. Camera stations currently used in the SMAAT project are shown in pink in Figure 5.

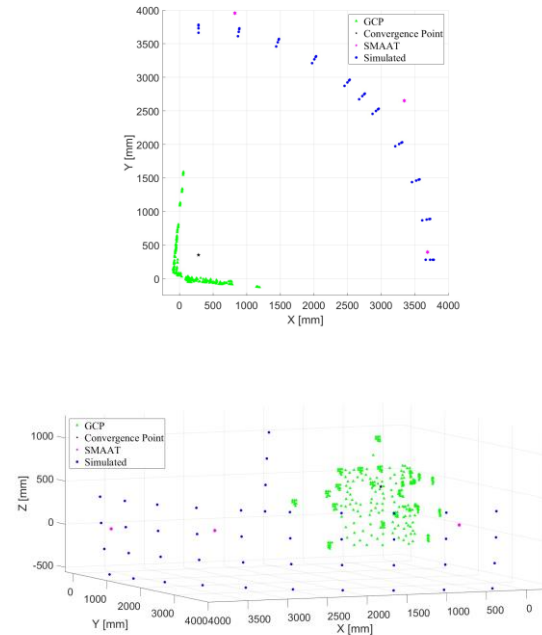


Figure 5: Simulated and SMAAT camera locations (blue) relative to a set convergence point (black) with a background GCP field (green) and SMAAT camera locations (purple).

Design Name	Horizontal Angle θ [°]	Elevation Angle ϵ [°]
D1	5	0
D2	5	5
D3	5	10
D4	5	15
D5	15	0
D6	15	5
D7	15	10
D8	15	15
D9	25	0
D10	25	5
D11	25	10
D12	25	15
D13	35	0
D14	35	5
D15	35	10
D16	35	15
D17	45	0
D18	45	5
D19	45	10
D20	45	15

Table 3: Network Designs for simulated networks.

3.3.1 Existing camera network: The simulated camera networks were compared to the existing one used for data capture for the SMAAT project. The locations of the existing camera network were determined by resecting the camera locations from observable GCPs in images taken from locations during SMAAT data capture.

3.4 Finding the optimal camera locations

Using the camera's EOPs, known object points and the predefined camera parameters (IOPs), collinearity equations were used to derive the image coordinates for each camera station. All points that did occur in all three images were removed.

Then, all network designs were processed through a least-squares adjustment (LSA). All image observations were equally weighted with an image coordinate standard deviation of 0.01mm (0.8313 pixels). The following outputs of the LSA were used for the further assessment of network uncertainty at a local level:

- Horizontal error ellipses representing horizontal uncertainty
- Standard deviations in height representing vertical uncertainty

For SMAAT, point accuracies and precisions must be high enough to facilitate appropriate analyses. Depth is very important for assessing motor speech control, such as the movement of the jaw, which is biomechanically characterised by six degrees of freedom, comprising jaw rotation and translation (Edward and Harris, 1990). Hence, the RMS values of the object space coordinates of the GCPs are an important indicator of how well networks perform for the SMAAT application. The reason is that the GCPs are equivalent to the facial markers, which will be tracked during the speech movement analysis. In this work, where GCPs are simulated and re-coordinated by the LSA, it is unrealistic to analyse the RMS outcomes because they are not a good interpretation of the true network performance. Hence, precision measures are the primary focus.

The error ellipses and standard deviation of the vertical axis of the object space coordinates of the GCPs were further investigated as they are better representations of the positional uncertainty of the adjusted point object space coordinates. An error ellipse around a point indicates the possible variances of the adjusted x- and y-coordinates at a 95% confidence level. Likewise, the standard deviation of the Z-axis is also expanded to a 95% confidence interval. The acceptance criteria are a small value of the semi-major axis (a), a small ratio of the semi-major and semi-minor axis (b) and low standard deviations in the Z-axis. The ratio indicates the shape of the error ellipse. A ratio closer to 1 means that the error ellipse has more of the shape of a circle, and subsequently, there is no directional dependency in the horizontal components.

Motivation	Aim	Measure
Higher depth precision for labial-facial speech assessment	Small positional uncertainty of object point coordinates	Small semi-major axis (a)
High precision for the extraction of facial landmarks for speech assessment	Ratio of semi-major (a) and semi-minor axis (b)	a/b close to "1"
	High precision in the vertical axis	Small σ_z

Table 4: Measures used for the assessments of the network designs.

4. Evaluation

4.1 Horizontal uncertainties

The point error ellipses were calculated for all GCPs per design and expanded to account for a 95% confidence level. As not all designs share the same GCPs, we analysed the mean values of the semi-major ellipse axis (a) and ellipse axis ratio (a/b) determined for each design.

The current SMAAT design is compared to the flat design (Figure 6, top graph) and vertically altered designs (Figure 6, bottom graph). For both the semi-major (a) and resultant ellipse ratio (a/b), the values in the simulated designs show an initial steep improvement. If the horizontal angle is equal to or larger than 70° , the graph follows a shallow improvement to the strongest design in D17/D20. This means, with one exception at D4, that each change in geometry is positively impacting the uncertainty of the GCP field. As expected, the SMAAT configuration performs well in both comparisons.

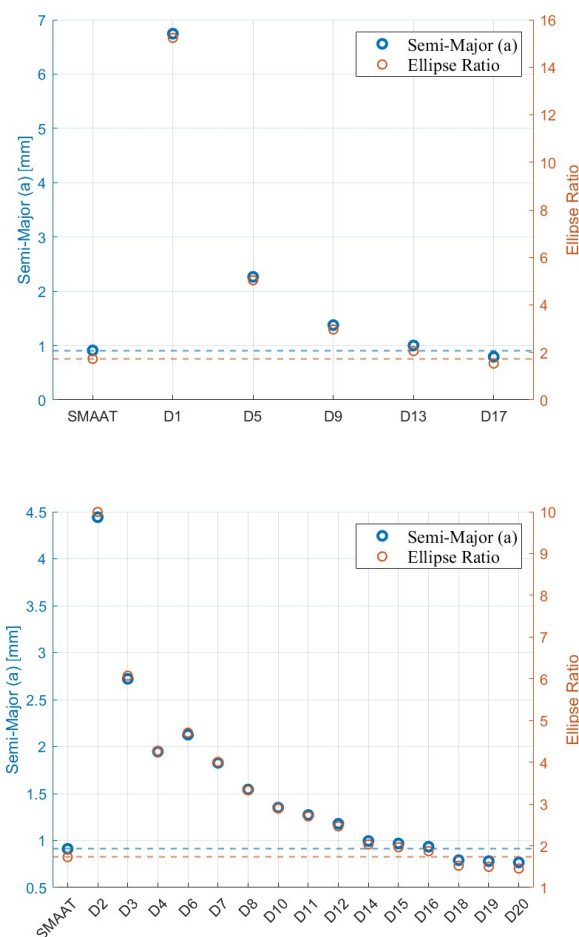


Figure 6: SMAAT design against flat designs (top) and vertically manipulated designs (bottom) for a and subsequent ratio value a/b .

Figure 7 shows the 95% error ellipses for a narrow design (D1, top) and the best design (D20, bottom). Both designs have a significant magnitude and ratio difference as already shown and discussed in Figure 6. All designs share the fact that the semi-major axis runs approximately parallel or toward the BMC, which confirms the challenge to estimate depth.

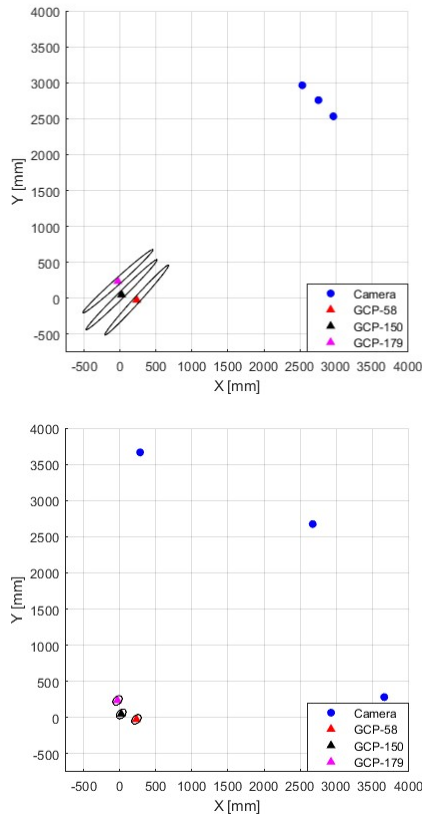


Figure 7: A reduced selection of 95% confidence interval error ellipses plots for narrow designs D1 (top) and the best-performing design D20 (bottom). All ellipses were scaled by a factor of 100.

For a more detailed analysis, Table 5 presents results for selected designs. D1 is a very narrow design ($\theta = 5^\circ$) with no vertical changes ($\varepsilon = 0^\circ$). D6 is a narrow design ($\theta = 15^\circ$) with a small vertical change ($\varepsilon = 5^\circ$). D11 is a normal design ($\theta = 25^\circ$) with a larger vertical change ($\varepsilon = 10^\circ$). And finally, D20 is the best-performing design with $\theta = 45^\circ$ and $\varepsilon = 15^\circ$. Full results are presented in Appendix A.

The largest mean value for the semi-major axis (a) is D1 (6.743 mm), but this rapidly improves with D6 (2.127 mm), followed by progressive improvement to D20 (0.767 mm). The current SMAAT design differs from D20 by only 0.145mm. Respectively, the mean a/b ratio for the current SMAAT design is only larger by 0.274 compared to the D20 design. Overall, the magnitude of the semi-major is constantly changing, with each geometrically stronger design having a relatively constant semi-minor.

Design	a [mm]	a/b
SMAAT	0.909	1.726
D1	6.743	15.240
D6	2.127	5.036
D11	1.272	2.705
D20	0.767	1.455

Table 5: A selection of designs and resultant 95% confidence interval error ellipse semi-major axis (a), and dual axis ratio (a/b).

4.2 Vertical uncertainty

The Z-axis (σ_z) precisions were calculated for all GCPs per design and expanded to account for a 95% confidence level. As not all designs share the same GCPs, the following precisions are reduced to mean values to provide more concise results.

Similar to Figure 6, the outcomes for the SMAAT design compared to flat and vertical changed locations are presented in Figure 8. All flat designs (Figure 8, top) have a vertical precision of less than 0.39 mm. However, beyond design D1, all preceding designs improve to 0.36 mm, regardless of flat or angled. The lowest vertical precision has design D17 ($\theta = 45^\circ$, $\varepsilon = 0^\circ$). The SMAAT design achieves 0.37mm. Meanwhile, a more complex scenario presents itself as vertical angles are introduced (Figure 8, bottom). The same initial steep improvement is followed by a slight improvement through to the final design. The lowest precision has a value larger than 0.4 mm, while the best precision is approximately 0.35 mm. Therefore, the inclusion of vertical convergence resulted in worse precision. Again, the SMAAT design is approximately in the middle with 0.37mm. It shares the same precision as the otherwise best-performing design D20.

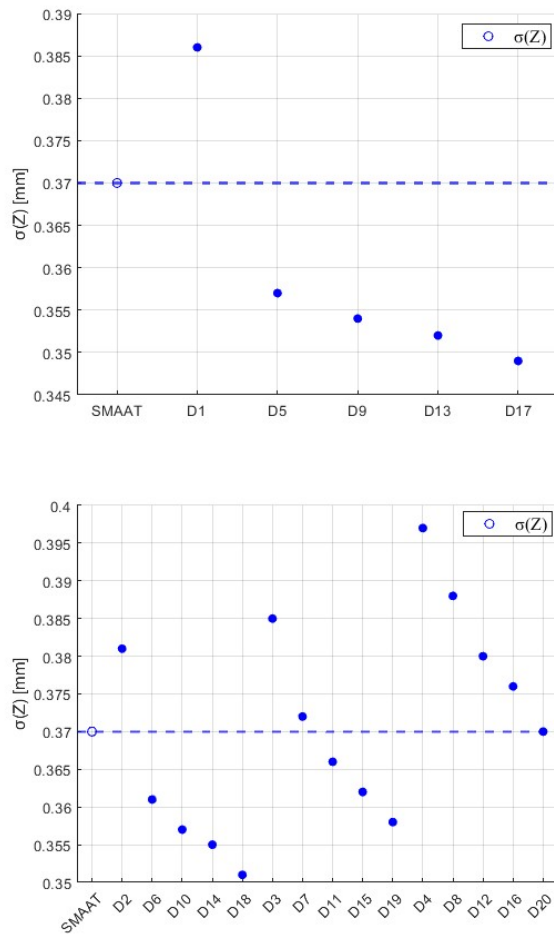


Figure 8: SMAAT design compared to flat designs (top) and all vertically manipulated designs (bottom) for σ_z .

The same selection of results as previously used is shown in Table 6; all results are presented in Appendix A. This table confirms the results presented in Figure 8, where increases in vertical convergence angle are not necessarily associated with better precision. For instance, design D20 is only superior to

design D1 and, in a best-case scenario, is equivalent to the SMAAT design. However, it should not be understated, as previously mentioned, that the general quantities are all low and consistently near 0.37 mm.

Design	σ_z [mm]
SMAAT	0.370
D1	0.386
D6	0.361
D11	0.366
D20	0.370

Table 6: A reduced selection of designs and resultant 95% confidence interval height precision (σ_z).

5. Discussion and conclusion

This paper's purpose was to test possible First-Order Designs suitable for measuring speech articulatory movements. We also aimed to compare the best designs to the existing configuration used in our SMAAT project. For this manuscript, we adopted a similar workflow as in Alsadika et al. (2012).

All network designs were simulated to converge at an approximated point close to the height of an adult head when seated for the speech assessment. Based on the constraints of the capture venue, various horizontally symmetrical designs were created with a maximum convergence of 90° comparable to the current SMAAT design. Vertical angle designs were also considered.

Our results suggest design D20 with 90° horizontal convergence and 30° vertical convergence is a superior design in terms of both the smallest error ellipse magnitude and the smallest ellipse axis ratio. Despite not achieving the most effective vertical precision, it outperforms over half of the other designs tested. However, in situations where D20 is unachievable, other designs may offer alternatives without significant sacrifice in precision. For example, D19 with a reduced vertical convergence of 20° only marginally reduces horizontal on the network, favouring a flatter design may prove advantageous. Results indicate that such designs can yield comparably small error ellipse quantities to designs with greater overall convergence (i.e. D20). However, the importance of maintaining horizontal convergence remains as the degradation of results occurs otherwise.

Contrary to theoretical expectations, vertical precision did not improve with progressive steps in angular geometry, rather, each increase in horizontal geometry yielded greater precision. The precision of design D17 ($\theta = 45^\circ, \varepsilon = 0^\circ$) can be attributed to the lack of vertical convergence, thereby minimising vertical uncertainty. Conversely, a purely flat design introduces errors in the horizontal plane.

Analysis of the presented results situates the existing SMAAT design within the top third to half of all tested designs. We found that whilst the strong horizontal geometry allows the horizontal uncertainty to be strongly controlled, the SMAAT network demonstrated weaker results in depth uncertainties; 19% less effective than D20 concerning semi-major axis and axis ratio values (a/b). Nonetheless, given the comparable axis ratio values (a/b) of 1.729 (SMAAT) versus 1.455 (D20), the difference may not significantly impact the assessment of speech movements.

Due to the comparable overall geometry, the SMAAT design closely resembled design D17. Whilst D17 exhibited similar points of uncertainties to D20, it proved more effective at minimising vertical precision. This can be attributed to the maximum vertical angle introduced to D20 that promotes more effective observations through greater angles of convergence. Our data suggests the SMAAT design, is comparable with the theoretical best (i.e., D20) design. For instance, design D20 was capable of a semi-major ellipse axis at a 95% confidence level of 0.767 mm. However, the SMAAT design was only 0.142 mm more at 0.909 mm. Likewise, the SMAAT ellipse ratio was 1.726 versus design D20 at 1.455, another relatively small difference. Finally, despite initial logical reasoning, the two designs share the same vertical precision of 0.370 mm. However, if the SMAAT workspace is factored in such that larger vertical convergences are less practical, design D18 is more appropriate while still maintaining strong horizontal geometry and satisfactory vertical convergence.

Compared to existing findings based on multiple calibrated camera positions, it can be concluded that the precision of the SMAAT design is suitable for measuring speech movements and comparable with previously reported markerless tracking systems using off-the-shelf cameras. For example, Feng and Max (2013) reported a vertical precision of 0.15 mm based on 60 frames per second footage and 3mm targets. If expanded to a corresponding 95% confidence interval, the difference compared to the SMAAT design is approximately 0.075 mm. While Feng and Max (2013) did calculate horizontal precision, they are separated into X and Y components rather than error ellipses. However, a simple and more effective design could be realised by maximising horizontal convergence and introducing a symmetrical vertical convergence as demonstrated in most of the simulated designs.

The SMAAT design, with its wide horizontal convergence and minor vertical convergence, was as good as the top third of tested designs. From a validation perspective, it can be concluded that the design is not only viable but also for continued use, allowing existing data to be used and all subsequent captures to be made with no changes or improvements based on the designs present in this paper.

Future research will focus on the investigation and implementation of the Second-Order Design principle (weighting). Additionally, a wide variety of semi-spherical designs or elliptical designs for 3 cameras (Ahmadabadian et al., 2014) will be investigated.

Acknowledgement

Research reported in this publication was supported by the Medical Research Future Fund under grant number 2016518, and completed as part of a doctoral research program.

References

Ahmadabadian A. H., Robson, S. R., & Boehm, J. (2014). Stereo-Imaging Network Design for Precise and Dense 3D Reconstruction. *The Photogrammetric Record*, 29(147), 317-336. doi.org/10.1111/phor.12076

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

Amini, A., S. (2017). Optimisation of Close Range Photogrammetry Network Design Applying Fuzzy Computations. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-4/W4, 31-36. doi.org/10.5194/isprs-archives-XLII-4-W4-31-2017

Barazzetti, L. (2017). Network Design in Close-Range Photogrammetry with Short Baseline Images. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, IV-2/W2, 17-23. doi.org/10.5194/isprs-annals-IV-2-W2-17-2017

Buschinelli, P., Salazar, J. D., Regner, D., Oliveira, D., Machado, M., Marcellino, G., Sales, D. C., Santos, J. M., Marinho, C. A., Stemmer, M. R., and Pinto, T. C. (2020). Targetless Photogrammetry Network Simulation for Inspection Planning in Oil and Gas Industry. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 1, 285-291. doi.org/10.5194/isprs-annals-V-1-2020-285-2020

Daniel, G. R., & McLeod, S. (2017). Children with speech sound disorders at school: Challenges for children, parents and teachers. *Australian Journal of Teacher Education (Online)*, 42(2), 81-101.

Dodd, B., 2014: Differential diagnosis of paediatric speech sound disorder. *Current Developmental Disorders Reports*, 1, 189-196. doi.org/10.1007/s40474-014-0017-3

Erdem, U. M., & Sclaroff, S. (2006). Automated camera layout to satisfy task-specific and floor plan-specific coverage requirements. *Computer Vision and Image Understanding*, 103(3), 156-169. doi.org/10.1016/j.cviu.2006.06.005

Fraser, C. S., 1984: Network Design Considerations for Non-Topographic Photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 50(8), 1115-1126.

Grafarend, E.W., & Sanso, F. (1985). Optimization and Design of Geodetic Networks. *Journal of Geodesy*, 59, 105-106. doi.org/10.1007/BF02519345

Jafari, D., Simmatis, L., Gaurin, D., Bouvier, L., Taati, B., & Yunusova, Y., 2023: 3D Video Tracking Technology in the Assessment of Orofacial Impairments in Neurological Disease: Clinical Validation. *Journal of Speech, Language, and Hearing Research*, 66(8S), 3151-3165. doi.org/10.1044/2023_JSLHR-22-00321

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical linguistics & phonetics*, 28(7-8), 508-531.

Olague, G., & Mohr, R. (2002). Optimal camera placement for accurate reconstruction. *Pattern Recognition*, 35(4), 927-944. doi.org/10.1016/S0031-3203(01)00076-0

Saadat, S. M., Samdzadegana, F., & Azizia, A. (2004). Camera Placement for Network Design in Vision Metrology Based on Fuzzy Inference System. *ISPRS Archives*. 35, 105-109.

Zhao, J., Cheung, S., & Nguyen, T. (2008). Optimal Camera Network Configurations for Visual Tagging. *IEEE Journal of Selected Topics in Signal Processing*, 2(4), 464-479. doi.org/10.1109/JSTSP.2008.2001430

Appendix A

Complete table outlining the resultant 95% confidence interval error ellipse semi-major axis (a), dual axis ratio (a/b), and vertical precision (σ_z).

Design	Mean a [mm]	a/b	σ_z [mm]
SMAAT	0.909	1.726	0.370
D1	6.743	15.240	0.386
D2	4.442	9.995	0.381
D3	2.722	6.073	0.385
D4	1.947	4.265	0.397
D5	2.267	5.036	0.357
D6	2.127	4.708	0.361
D7	1.827	4.008	0.372
D8	1.543	3.329	0.388
D9	1.378	2.965	0.354
D10	1.350	2.895	0.357
D11	1.272	2.705	0.366
D12	1.178	2.467	0.380
D13	1.003	2.057	0.352
D14	0.995	2.033	0.355
D15	0.969	1.966	0.362
D16	0.934	1.871	0.376
D17	0.795	1.531	0.349
D18	0.791	1.521	0.351
D19	0.781	1.494	0.358
D20	0.767	1.455	0.370