

Evaluation of properties and usability of virtual reality interaction techniques in craniomaxillofacial computer-assisted surgical simulation

Helena Mehtonen

Medical Imaging Centre, Department of Radiology, Tampere University Hospital, 33520 Tampere, Finland
Faculty of Medicine and Health Technology, Tampere University, 33520 Tampere, Finland
Faculty of Information Technology and Communication Sciences, Tampere University, 33014 Tampere, Finland
helena.mehtonen@pirha.fi

Jorma Järnstedt

Medical Imaging Centre, Department of Radiology, Tampere University Hospital, 33520 Tampere, Finland
Faculty of Medicine and Health Technology, Tampere University, 33520 Tampere, Finland
jorma.jarntstedt@pirha.fi

Jari Kangas

Faculty of Information Technology and Communication Sciences, Tampere University, 33014 Tampere, Finland
jari.a.kangas@tuni.fi

Sriram Kishore Kumar

Faculty of Information Technology and Communication Sciences, Tampere University, 33014 Tampere, Finland
sriramkumarkishorekumar@tuni.fi

Irina Rinta-Kiikka

Medical Imaging Centre, Department of Radiology, Tampere University Hospital, 33520 Tampere, Finland
Faculty of Medicine and Health Technology, Tampere University, 33520 Tampere, Finland
irina.rinta-kiikka@pirha.fi

Roope Raisamo

Faculty of Information Technology and Communication Sciences, Tampere University, 33014 Tampere, Finland
roope.raisamo@tuni.fi

ABSTRACT

Although there are promising results from the use of virtual reality (VR) in the craniomaxillofacial field, there is still a need to validate the usability and properties of the VR environment and interaction techniques. The aim of this study was to evaluate the suitability of VR interaction methods for craniomaxillofacial computer-assisted surgical simulation (CASS) and to identify possible areas for improvement.

Four VR interaction conditions were compared quantitatively and qualitatively: Hand, Mouse, Pen and, Controller. Four oral and maxillofacial radiologists performed a VR marking task on skull stereolithography models. Quantitative measures included accuracy, completion time, number of grasps and development. Qualitative attributes were easiness, efficiency, physical effort, accuracy, and naturalness.

Mouse (1,51 mm) and Controller (1,73 mm) were the most accurate, Pen was the close third (2,06 mm), while Hand (4,52 mm) scored poorly. Mouse was slower and more burdensome than the other conditions. The accuracy of Pen and the completion times of

Hand, Pen and Mouse improved over time. The usability of Controller (1,50) was rated best in Likert scale (1-5), with Pen (1,75) the close second. Mouse (3,00) and Hand (3,57) were inferior, and overall, Hand was the least preferred.

Controller and Mouse achieved acceptable accuracy for craniomaxillofacial CASS. The usability of Controller was also rated highest, and it was the preferred choice of the radiologists. The combination of Mouse and VR was unnatural and cumbersome. To achieve an acceptable level of accuracy for Hand, hand tracking technology needs to be significantly improved.

CCS CONCEPTS

• Human computer interaction (HCI); • Life and medical sciences; • Modeling and simulation;

KEYWORDS

Computer-assisted surgical simulation, Virtual reality, Interaction techniques, Craniomaxillofacial radiology, Craniomaxillofacial surgery, Cone beam computed tomography

ACM Reference Format:

Helena Mehtonen, Jorma Järnstedt, Jari Kangas, Sriram Kishore Kumar, Irina Rinta-Kiikka, and Roope Raisamo. 2023. Evaluation of properties and usability of virtual reality interaction techniques in craniomaxillofacial computer-assisted surgical simulation. In *OzCHI 2023 (OzCHI 2023)*, December 02–06, 2023, Wellington, New Zealand. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3638380.3638433>

1 INTRODUCTION AND BACKGROUND

In the craniomaxillofacial (CMF) region, technological achievements enable a shift from conventional planning to three-dimensional (3D) patient-specific computer-assisted surgical simulation (CASS), as the quality and amount of information in medical imaging scans have increased continuously [1–3]. However, the visualisation means have not developed at the same pace. Currently, radiological interpretation and CASS are still performed using a traditional two-dimensional (2D) computer interface with mouse and keyboard, even though the scans are often, and the anatomical region of interest always, three-dimensional. This lack of true 3D immersion requires radiologists to have a profound spatial understanding and makes the interpretation and CASS time-consuming and error-prone [4].

Craniomaxillofacial CASSs are used for orthognathic surgery [5, 6], resection and reconstructive cases such as tumours, traumas, and deformities [7–11] and planning other osteotomy guides [12]. CASSs offer the ability to simulate different surgical procedures preoperatively, helping to define the most appropriate surgical approach [13] and providing both better efficiency [14] and accuracy [6] compared to conventional planning. Accurate planning of CMF procedures is essential to identify major risks and avoid complications, e.g., bleeding and nerve injury [15–20].

Virtual reality (VR), a true 3D environment, has been shown to provide an advantage over 2D visualisation for perception and comprehension [21–23]. However, VR has yet to be validated into CMF CASS. There is only a limited number of orthognathic surgery studies focusing on VR environment [24, 25] and only one anatomical landmarking study which concentrated on interaction methods and/or user experience (UX) showing a preference for the stylus over the mouse and hands [26]. The results are promising but sparse, and thorough testing of interaction methods and their characteristics is necessary.

Computer-aided surgical simulation is comparable to any domain where computer-aided design (CAD) is used, and three-dimensional perception is required. Thus, there should be many potential comparable domains. However, only a few comparative studies focusing on multiple interaction methods across domains have been reported. An industrial virtual reality study compared controllers and gloves in the field of low-cost automation. The usability of the gloves was rated higher than that of the controllers [27]. A computer science study compared a mouse, a controller and a 3D pen-like device in a VR and AR pointing task. The results showed that the pen outperformed other VR controllers in terms of efficiency and was comparable to the mouse. Participants also preferred the pen to the controllers [28]. In architecture, virtual reality and controllers were compared with the traditional mouse in a 2D interface. The results showed that interaction with virtual reality and controllers was more natural than with a mouse in a 2D environment [29].

The aim of this study was to medically validate the suitability of four VR interaction methods for CMF CASS: hand, mouse, Logitech VR Ink stylus, and Valve Index Controller. CASS and VR were combined for a marking task on facial bone stereolithography (STL) models based on cone beam computed tomography (CBCT). Results on accuracy, completion time, number of grasps interpreting physical workload, and UX were compared. The experiment was divided

into five test sessions over two months to examine the development of the above attributes.

2 MATERIALS AND METHODS

2.1 Participants

Four oral and maxillofacial (OMF) radiologists with experience in CASS but not in VR were recruited. Their experience in dentistry and radiology ranged from 7 to 35 years and in CASS from 5 to 20 years.

2.2 Medical data

50 stereolithography skull models were segmented from preoperative facial CBCTs of orthognathic surgery patients using Materialise ProPlan CMF 3.0.1. The CBCTs were scanned with Scanora 3Dx by Soredex or Promax Mid by Planmeca at the Department of Cranio- and Dentomaxillofacial Radiology, Tampere University Hospital, Finland. Voxel sizes varied between 0.2–0.4 mm.

2.3 Legality of data use

This study was based on a retrospective registry data set and did not involve human experimentation or the use of human tissue samples. No patients were imaged for this study. The study followed Finnish legislation, the Medical Research Act (488/1999) and the Act on Secondary Use of Health and Social Data (552/2019), and the European General Data Protection Regulation 216/679. The use of data was approved by the research director of Tampere University Hospital, Finland, on 1 October 2019 (vote number R20558).

2.4 Data preparation

Five 2 mm diameter spheres were placed higgledy-piggledy on the surface of each skull with the Valve Index Controller in the Unity environment (Figure 1).

2.5 Interaction conditions

Four dominant hand interaction techniques were used: Hand, Mouse, Pen, and Controller. The instruments and hands were visible in the VR space as animated forms. The devices ray-casted a transparent beam with a spherical tip similar in diameter to predefined spheres (Figure 2).

Participants had only one chance to make each mark; correction, erasure, scaling, or zooming were disabled, as the emphasis was on an equal and straightforward comparison of conditions.

The skull model was moved using Valve Index Controller and non-dominant hand. The controller+grasp method was used because of its suitability for a natural and accurate grasping [30].

2.5.1 Hand. A poke is used to indicate and touch surfaces, so this gesture was chosen for Hand condition. The marking occurred slightly before contact, otherwise the participant's finger would have covered the beam tip and the pre-defined sphere.

2.5.2 Mouse. For the Mouse condition, a conventional mouse resting on a table was chosen. Movements to forward and backward transferred the projected beam up and down and shift to left and right correlated in both environments. Marking was done by pressing the left mouse button.

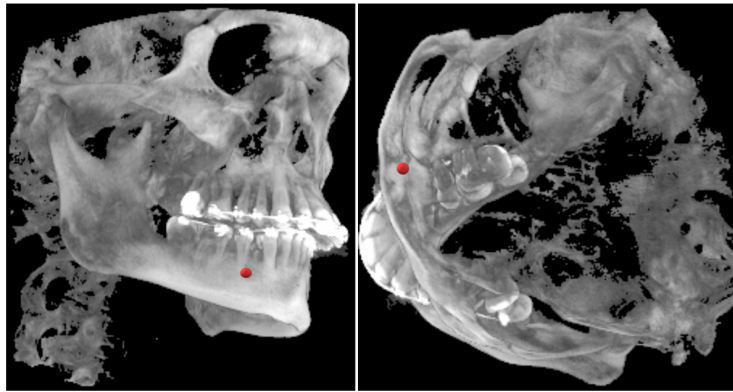


Figure 1: Two examples of the higgledy-piggledy placed 2 mm diameter spheres on the skull



Figure 2: Experiment room, VR environment with training cube, and interaction methods from left to right: Hand, Mouse, Pen and, Controller

2.5.3 Pen. The Logitech VR Ink Stylus, which has a physical similarity to a traditional pen, was chosen for Pen condition. Marking happened by tapping the primary button with index finger.

2.5.4 Controller. The Valve Index Controller was chosen for the Controller condition. It has hand and finger tracking, and a tight strap that allows it to be held passively. Marking was done by pressing the trigger button with index finger.

2.6 VR environment and experiment room

The VR environment was created using the Unity 3D software development system and the Varjo VR-2 Pro head-mounted display (HMD). It included an open workspace, a button, and a progress panel. Hand was tracked by the Varjo HMD and Pen and Controller by the HTC Vive base stations. The VR environment was set up in an office room containing an office desk and an office chair with armrests and headrest.

2.7 Procedure and session details

In the first of the five experimental sessions, the radiologists were introduced to the study protocol and the environment. Before each condition, the devices were introduced, and the participants were asked to mark the pre-placed spheres as accurately as possible. In the following four sessions, the introduction was briefly repeated. One session included all four conditions, ten skulls per method, and lasted approximately 60 minutes.

The order of the conditions was counterbalanced using the Latin square method. Each condition contained ten skull models in random order, and every skull contained five pre-placed spheres. In each condition, the first skull was preceded by a cube with five pre-marked points on which the participant practiced the marking task.

Participants moved to the next model by pressing the button on the left side of the VR room. On the participant's right, a progress panel showed the current number of skulls per method (e.g., 8/10) and the number of points on the skull marked out of a total of 5 (e.g., 2/5).

Table 1: The statements to evaluate the qualitative attributes of interaction conditions

Attribute	Statement	Rating scale
Easiness	The interaction with this method was	1 easy – 5 difficult
Efficiency	The interaction with this method was	1 efficient – 5 inefficient
Physical effort	The interaction with this method was physically	1 effortless – 5 wearing
Accuracy	The interaction with this method was	1 accurate – 5 inaccurate
Naturalness	The interaction with this method was	1 natural – 5 unnatural

After each condition radiologists filled a usability questionnaire and at the end of each session, they ranked the conditions from best to worst and gave reasons for the rating. In total, each participant marked 200 skulls and 1000 points. Quantitative measures were recorded automatically, and qualitative responses were collected via questionnaires.

3 MEASUREMENTS

3.1 Quantitative

The primary measure was accuracy, defined as the distance in millimetres between the centres of the pre-marked and participant-marked spheres. Other measures were completion time, number of non-dominant hand grasps representing physical effort, and development. Time was recorded in seconds, starting from the first grasp, and stopping after all five points had been marked and the button had been pressed. A grasp occurred when participant grabbed and released the model. Development was defined as a change between sessions 1 and 5.

3.1.1 Statistical analysis for quantitative values. Median values and Monte Carlo permutation test were used for quantitative statistical analyses because the data contained some outliers caused by technical malfunction or human error. In all permutation tests, an observed value of a measure is compared to a distribution of measures generated by resampling multiple sample permutations, assuming no difference between the sample sets (null hypothesis). The relevant p-value is then given by the proportion of values in the distribution that are more extreme than or equal to the observed value [31–34].

Possible statistically significant differences between and within the four conditions were analyzed. To obtain the distribution of measurements assuming no difference, samples were randomly interchanged between conditions, generating 10,000 permutations to be measured. To counteract the multiple comparisons problem, the Bonferroni correction was used.

3.2 Qualitative

A 5-point Likert scale (1-5) statement questionnaire, filled in after each condition, was used for the usability evaluation, with opposite adjectives representing the extremities. The attributes were easiness, efficiency, physical effort, accuracy, and naturalness (Table 1). Radiologists also listed at least one positive and one negative comment about each condition. At the end of the session, conditions were rated from most preferred (1) to least preferred (4). Comments about interaction with the non-dominant hand were also collected.

3.2.1 Qualitative values. Usability ratings were considered as opinions and the focus was on collecting comments. The adjective scale was changed to a numerical one where a score of 1 correlated with a positive rating and a score of 5 with a negative rating as shown in Table 1. Qualitative results were expressed as mean scores.

4 RESULTS

4.1 Quantitative

The results are shown in Figure 3 and Tables 2 and 3. For accuracy, time, and number of grasps the Bonferroni correction of $0.05/6 \sim 0.0083$ was used. The Bonferroni correction for development was $0.05/1 = 0.05$ as development took place within each technique and thus there was only one condition in the permutation test.

4.1.1 Accuracy. For accuracy, the nominator was per point (1000 values per condition). A statistically significant difference was found between all interaction methods. The results from least accurate to most accurate in millimetres was as follows: Hand (4.52), Pen (2.06), Controller (1.73), and Mouse (1.51). The p-values for six different comparisons were: Controller + Hand <0.0000 , Controller + Mouse 0.0006 , Controller + Pen 0.0001 , Hand + Mouse <0.0000 , Hand + Pen <0.0000 , Mouse + Pen <0.0000 .

Medically accepted accuracy was also considered. 2 mm has been defined as the clinically accepted limit between the virtual plan and the postoperative outcome and has been previously used to verify the accuracy of 2D CASSs [6, 7, 35–37]. Within this limit, the results showed acceptability for Controller and Mouse.

4.1.2 Time. For completion time, the nominator was per skull (200 values per condition) and the results were in seconds. Mouse (40,88) was significantly slower than other conditions: Hand (32,72), Controller (31,99), and Pen (30,48). The p-values for six different comparisons were: Controller + Hand 0.7511 , Controller + Mouse <0.0000 , Controller + Pen 0.3231 , Hand + Mouse <0.0000 , Hand + Pen 0.1074 , Mouse + Pen <0.0000 .

4.1.3 Number of grasps. For number of grasps, the nominator was per skull (200 values per condition). Mouse required more grasps than the other methods: Mouse (18,00), Controller (14,50), Pen (14,00) and, Hand (14,00). The p-values: Controller + Hand 0.9244 , Controller + Mouse <0.0000 , Controller + Pen 1.0000 , Hand + Mouse <0.0000 , Hand + Pen 1.0000 , Mouse + Pen <0.0000 .

4.2 Quantitative development

Development was defined as the change in the radiologists' combined scores for accuracy, completion time and number of grasps between sessions 1 and 5. The nominator for accuracy was per

Table 2: The quantitative results for accuracy, completion time and number of grasps

Conditions	Accuracy (mm) per point	Time (s) per skull	Grasps (number) per skull
Hand	4.52	32.72	14.00
Mouse	1.51	40.88	18.00
Pen	2.06	30.48	14.00
Controller	1.73	31.99	14.50

Table 3: The p-values of compared conditions for accuracy, completion time and number of grasps, with Bonferroni correction for p-value limit $0.05/6 \sim 0.0083$

Compared conditions	Accuracy (per point)	Time (per skull)	Number of grasps (per skull)
Controller - Hand	<0.0000	0.7511	0.9244
Controller - Mouse	0.0006	<0.0000	<0.0000
Controller - Pen	0.0001	0.3231	1.0000
Hand - Mouse	<0.0000	<0.0000	<0.0000
Hand - Pen	<0.0000	0.1074	1.0000
Mouse - Pen	0.0004	<0.0000	<0.0000

Table 4: The p-values of development for accuracy, completion time and number of grasps within each condition, with Bonferroni correction for p-value limit $0.05/1 = 0.05$

Conditions	Accuracy	Time	Grasps
Hand	0.2123	<0.0001	0.1881
Mouse	0.5050	<0.0001	0.6670
Pen	0.0491	0.0342	0.1862
Controller	0.5652	0.2127	0.1280

mark (200 values per session) and for completion time and number of grasps per skull (40 values per session).

The development within different attributes is shown in Figure 4 and Table 4. For accuracy, an improving trend and a significant difference was found within Pen (p-value 0.0491). The p-values for the other conditions were: Hand 0.2123, Mouse 0.5050, and Controller 0.5652. For completion time, there was a significant improvement within Hand (p-value < 0.0000), Mouse (p-value < 0.0000) and Pen (0.0342). The p-value for Controller was 0.2117. There were no significant differences for grasping, for which the p-values were: Controller 0.1280, Pen 0.1862, Hand 0.1881, and Mouse 0.6670.

4.3 Qualitative

The mean values of the usability questionnaires are shown in Figure 5. After some fluctuation, Controller was the preferred method or as good as the others for all attributes, with accuracy improving towards the end. Pen was either in second or tied for first and its performance was close to that of Controller. The usability of Mouse and Hand was inferior to Pen and Controller, except for accuracy, with which Mouse shared the first place. Mouse was initially the least liked, but the ranking gradually improved, and was eventually preferred to Hand, except for naturalness. Overall, Hand was the least liked instrument, and particularly its accuracy was judged to be extremely poor.

4.4 Ranking

The results are shown in Figure 6 and mean values in Table 5. Controller was the most preferred condition, although its ranking decreased slightly after the beginning. Pen was second and the difference between Pen and Controller decreased towards the end. Mouse and Hand were less preferred. Hand was well received at the beginning, but its rating dropped towards the end. Conversely, Mouse was strongly disliked at first, but later bypassed Hand and came in third.

4.5 Comments and notions

4.5.1 Hand. Hand initially scored well because of its instrument-free status. It was logical and intuitive to use, and pointing was natural way to interact. However, radiologists became frustrated with poor accuracy and problems with tracking. The correct HMD-tracked position was difficult to remember, and the finger became physically tired with continued extension. Poor tracking and anticipatory marking functionality made it difficult to judge when the marking would occur. The progressive increase in frustration was commented on as a reason for the decline in rating.

4.5.2 Mouse. Familiarity and accuracy were positive features of Mouse. The ergonomics of the dominant hand were appreciated as the hand resting on the table eliminated the natural instability. However, it took time to learn to use the mouse in VR and

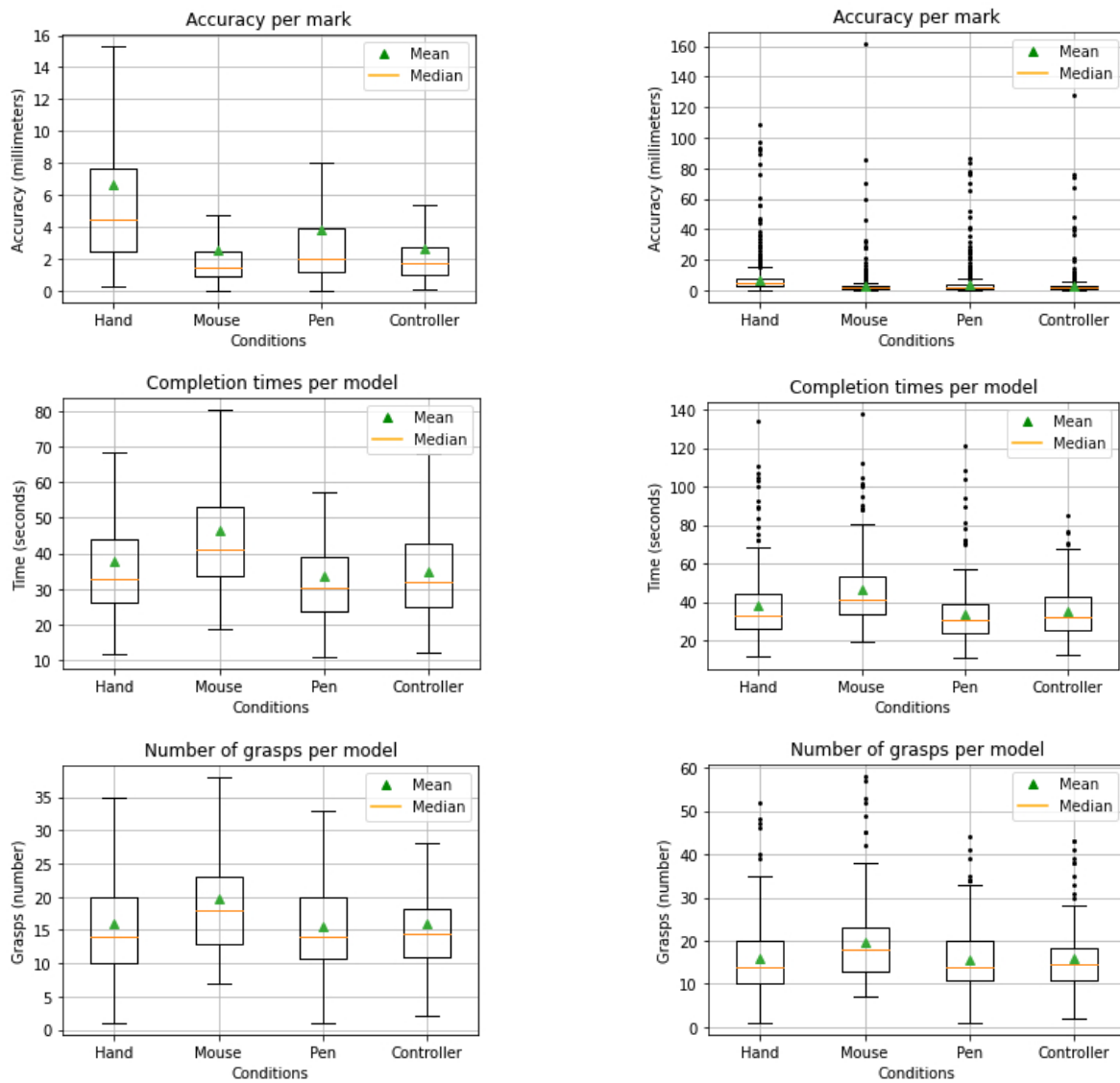


Figure 3: The mean and median values of conditions for accuracy, completion time and number of grasps, with and without outliers.

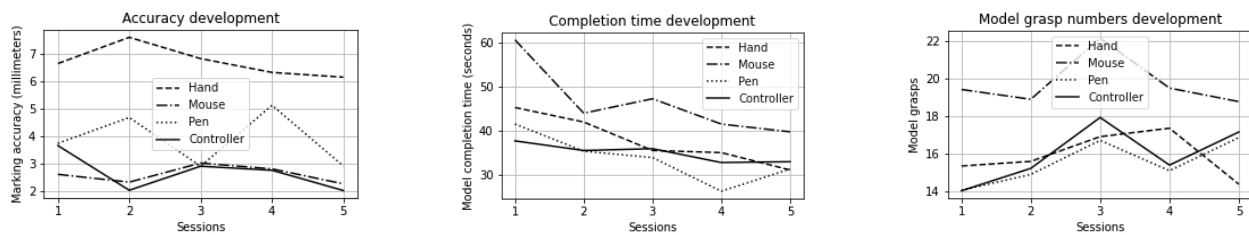


Figure 4: Development for accuracy, completion time and number of grasps within each condition

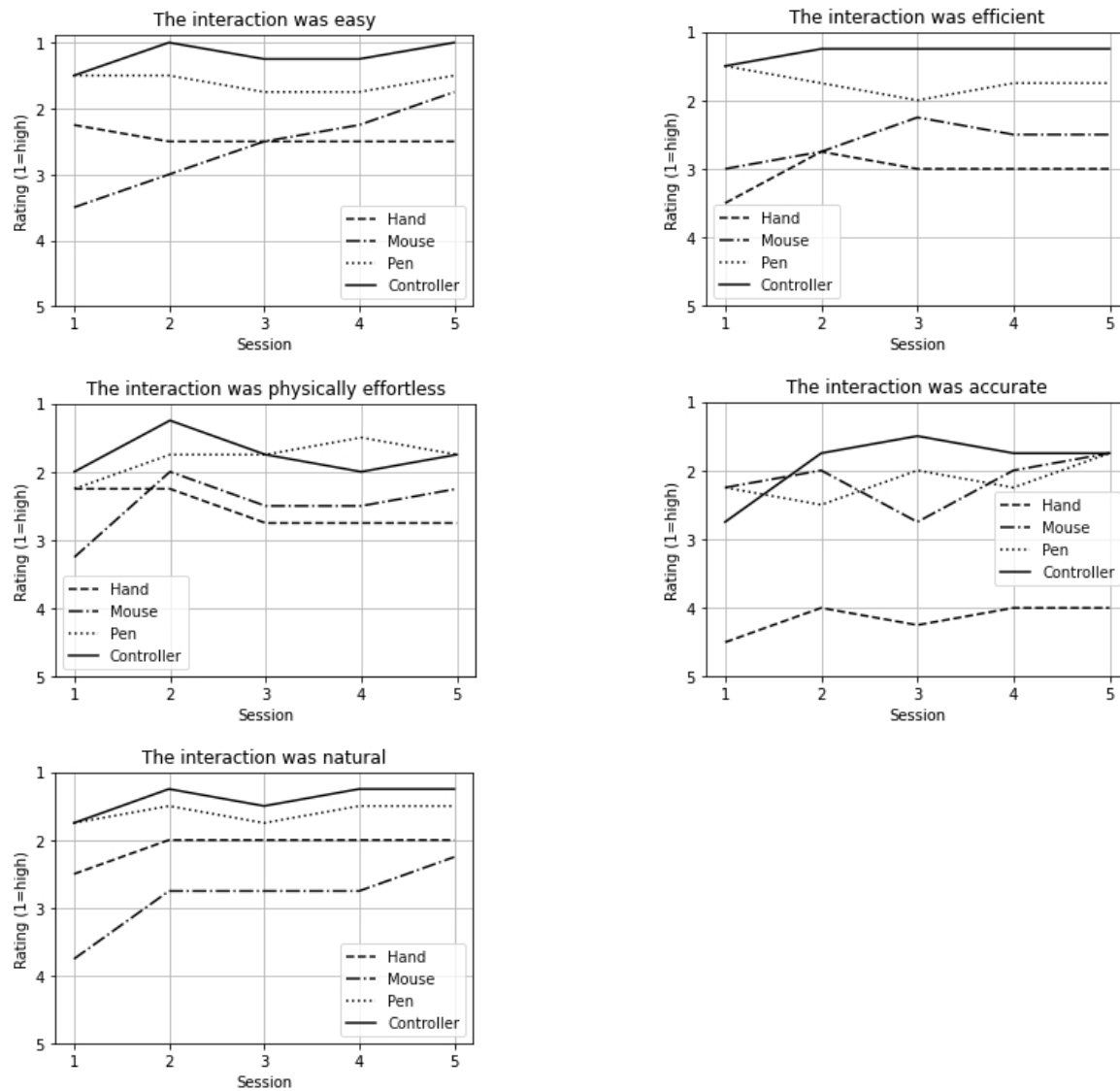


Figure 5: Ratings of qualitative statements in 5-step mean Likert scale where 1 = high rating, 5 = low rating

Table 5: The ranking of conditions per sessions, mean values, 1 = high, 4 = low

Condition	Session 1	Session 2	Session 3	Session 4	Session 5
Hand	2.50	3.50	3.50	3.75	3.75
Mouse	3.75	3.50	3.50	3.00	3.00
Pen	2.00	2.00	1.50	1.75	1.75
Controller	1.75	1.00	1.50	1.50	1.50

the restriction of the dominant hand in 2D led to unnaturalness and physical fatigue of the non-dominant hand. The movements of the non-dominant hand were increased to compensate for the restricted dominant hand. In addition, some parts of the skull, e.g., the posterior part of the orbita, were difficult to reach due to the

lack of 3D range. Finally, the need to remain close to table caused radiologists to occasionally bang their non-dominant hand to the table.

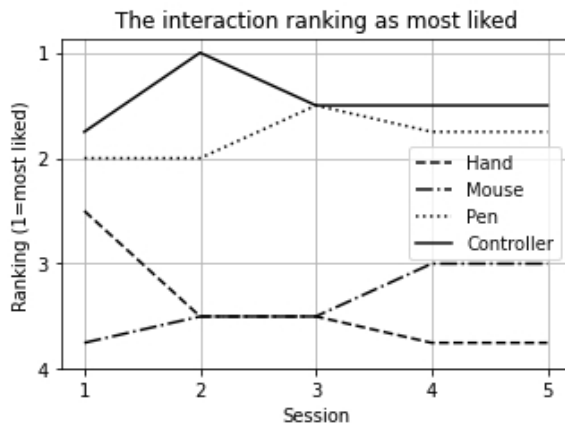


Figure 6: Ranking development of conditions from session 1 to session 5

4.5.3 Pen. The Logitech VR Ink was liked as a light and familiar tool. It was logical and easy to use. However, the buttons were a disappointment. They did not provide proper tactile feedback which led radiologists to report increased inaccuracy as they felt uncertain about when the marking would occur. Further inaccuracy was reported due to natural tremor.

4.5.4 Controller. Controller was natural, comfortable, ergonomic, and easy to learn and use. Accuracy was appreciated, although participants repeatedly commented that the controller was too sensitive, replicating the natural tremor of the hand and leading to inaccuracy. This was the main reason for Controller's decline in rating. One radiologist compensated for the tremor by stabilising the dominant hand with the non-dominant hand. Another disadvantage was the big size of the controller. It was suspected that the size could lead to physical fatigue. This, combined with sensitivity, could significantly reduce accuracy with prolonged use.

4.5.5 Other comments. Radiologists indicated that translating and rotating the model by holding the non-dominant hand in extension resulted in physical fatigue. This led to a request for a zoom in/out function. For Controller and Pen, there were also requests for better elbow support and built-in motion stabilisation.

5 DISCUSSION

In recent years, there has been a growing interest in the use of VR in craniomaxillofacial field, and preliminary studies show promising results. However, there is still a need to validate the usability and properties of the VR environment and the interaction techniques. The main requirements of VR instruments for CMF CASS include accuracy, speed, and ergonomics. The UX must also be satisfactory.

Four VR devices were included in this study. Two were commercially produced: a Valve Index Controller and a Logitech Ink Stylus. The other two were a standard 2D mouse and a hand immersed in VR. Four OMF radiologists took part in the test protocols. Although they all had experience with CASS, the VR environment was new or almost new to them. Therefore, besides the comparison, the

development was analysed to see if increasing experience would change the feasibility of each condition.

This study was the first to combine the comparison of technical performance and usability of different VR interaction methods in CMF CASS. The previous anatomical landmarking study [26], based on a limited number of data, showed parallel results by preferring the stylus over the mouse and hand, but lacked the controller and medical experts.

Our primary quantitative measure, accuracy, showed acceptable results for Controller and Mouse. However, the latter was significantly slower and required more grasps than all the other conditions. With Mouse, the improvement in accuracy was limited due to the restricted range. The limited 2D range also seemed to explain the significant difference in the number of grasps, slower completion time, and why Mouse was reported to cause physical fatigue. With Controller, the accuracy and completion time did not improve due to the tremor and its compensation with reduced speed. This indicates a need for built-in motion stabilization and better physical hand support. The differences in completion time and number of grasps between Hand, Pen and Controller were not statistically significant.

The current level of hand tracking technology for Hand proved to be insufficient. The fingertip obscured the small targets, and the limited spatial coverage of the HMD's cameras made the positioning of the finger difficult. For Hand, the radiologists supported a separate marking command, e.g., a voice command. On the other hand, the problems with handtracking could be compensated by using a glove which was the preferred interaction method in an industrial study on low-cost automation [27].

The radiologists were disappointed with the physical features of Pen. This was opposite in contrast to a computational science study where the pen performed better and was also preferred to the controller [28]. The tactile feedback was faint, the finger tended to slip off the button, which was difficult to rediscover with the HMD on, and Pen also replicated natural tremor. This suggests a need for a stylus with more distinct buttons, haptic feedback, and built-in motion stabilization.

During development, the accuracy of Pen improved, and completion times of Hand, Mouse and Pen shortened. The positive trend for Mouse and Pen seemed to be due to learning, but the shorter completion time for Hand seemed to relate to frustrated radiologists rushing through the condition, causing further inaccuracy.

Quantitative and qualitative results were parallel although the participants initially reported a slight and only recognized bias towards Hand and Mouse because their familiarity: Controller achieved good quantitative results and was the most preferred, while Hand was the worst in both performance and preference, with the Mouse and Pen in between. In architecture, the virtual reality environment and controllers were also preferred to the traditional mouse in a 2D interface [29].

Our results are not limited to CMF CASS but can be generalised to various types of surgical plannings in other parts of the body. The aim of VR CASS is to facilitate understanding of three-dimensional structures, further improve surgical accuracy, reduce operating time and complications, and lead to more cost-effective treatment. Virtual reality can also be used to teach anatomy, pathology, and surgery. Furthermore, it enables real-time teaching, consultation,

and surgical planning between medical specialists and/or students in the same virtual environment, wherever they are in the world.

Our study had some technical limitations. Only one device per interaction method type was used, representing a technical implementation from a single manufacturer. A similar limitation is the use of one HMD. The accuracy of hand tracking may vary between different manufacturers, which may have led to differences in the results of Hand. The size of the headset was also an issue, as one participant was unable to fit display glasses under the HMD, and prescription lens inserts were not available. This highlights the need for easily accessible individual correction lenses or built-in diopter correction.

The limitation of the software was that there were no correction or fine-tuning functions. Therefore, the marking task did not fully simulate a real computer-assisted surgical simulation. The use of surface models, combined with the lack of radiological multi-sectional view, caused problems with reach, as some spheres were obscured by other anatomical structures.

Because of these limitations and the overall scarcity of worldwide research, further studies are needed to provide precise recommendations or guidelines on the most appropriate choice of interaction tools for CMF CASS. However, our results suggest that operating in the VR could be suitable for CASS if interaction methods and ergonomics are further developed and validated. To improve accuracy, motion stabilization and erasure and correction functions are needed. For ergonomics, there is a demand for a zoom-in/zoom-out function, as smaller models require a smaller range of motion. Ergonomics can be further improved by designing a more suitable arm support for VR instruments operated in six-degrees of freedom.

Our next study will simulate a realistic CMF CASS VR environment with a multi-section radiological slice view and a more complex interaction task, optimised by the results and identified limitations of this study. We will compare a traditional two-dimensional computer interface with a VR, which is crucial to validate the technical performance and usability of the VR method, including user satisfaction and cognitive load.

6 LIMITATIONS

Our results are specific only for software system, headset, and interaction techniques used in this study, even if they indicate potential for a larger group of similar devices. It should also be noted that this kind of marking task does not describe all the functions required in CASS. Other functions need to be addressed in further work.

7 CONCLUSIONS

Two VR interaction methods, Controller and Mouse, achieved acceptable accuracy for craniomaxillofacial CASS, though the combination of Mouse and VR was unnatural and cumbersome. The usability of Controller was rated highest, and it was the preferred choice of the CMF radiologists. To reach an acceptable level of tracking for Hand, hand tracking technology needs to be clearly improved.

ACKNOWLEDGMENTS

This study was supported by Business Finland under project “Digital and Physical Immersion in Radiology and Surgery” (Grant no 930/31/2019) and by Academy of Finland under project “Explainable AI Technologies for Segmenting 3D Imaging Data” (Grant no 345448).

REFERENCES

- [1] Mariusz Wrzosek, Zachary Peacock, Amir Laviv, BR Goldwaser, R Ortiz, CM Resnick, MJ Troulis, and LB Kaban. 2016. Comparison of time required for traditional versus virtual orthognathic surgery treatment planning. *International journal of oral and maxillofacial surgery* 45, 9 (2016), 1065–1069. DOI:https://doi.org/10.1016/j.ijom.2016.03.012
- [2] Rebecca Smith-Bindman, Diana L. Miglioretti, Eric Johnson, Choonsik Lee, Heather Spencer Feigelson, Michael Flynn, Robert T. Greenlee, Randell L. Kruger, Mark C. Hornbrook, Douglas Roblin, Leif I. Solberg, Nicholas Vaneman, Sheila Weinmann, and Andrew E. Williams. 2012. Use of Diagnostic Imaging Studies and Associated Radiation Exposure for Patients Enrolled in Large Integrated Health Care Systems, 1996–2010. *JAMA* 307, 22 (June 2012). DOI:https://doi.org/10.1001/jama.2012.5960
- [3] Georgios Sakas. 2002. Trends in medical imaging: from 2D to 3D. *Computers & Graphics* 26, 4 (August 2002), 577–587. DOI:https://doi.org/10.1016/S0097-8493(02)00103-6
- [4] Daniel Mendes, Fabio Caputo, Andrea Giachetti, Alfredo Ferreira, and J. Jorge. 2019. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum* 38, 1 (February 2019), 21–45. DOI:https://doi.org/10.1111/cgf.13390
- [5] Hsiu-Hsia Lin and Lun-Jou Lo. 2015. Three-dimensional computer-assisted surgical simulation and intraoperative navigation in orthognathic surgery: A literature review. *Journal of the Formosan Medical Association* 114, 4 (April 2015), 300–307. DOI:https://doi.org/10.1016/j.jfma.2015.01.017
- [6] Ali Alkhayer, József Pifkó, Carsten Lippold, and Emil Segatto. 2020. Accuracy of virtual planning in orthognathic surgery: a systematic review. *Head Face Med* 16, 1 (December 2020), 34. DOI:https://doi.org/10.1186/s13005-020-00250-2
- [7] James J. Xia, Jaime Gateno, John F. Teichgraber, Andrew M. Christensen, Robert E. Lasky, Jeremy J. Lemoine, and Michael A.K. Liebschner. 2007. Accuracy of the Computer-Aided Surgical Simulation (CASS) System in the Treatment of Patients with Complex Craniomaxillofacial Deformity: A Pilot Study. *Journal of Oral and Maxillofacial Surgery* 65, 2 (February 2007), 248–254. DOI:https://doi.org/10.1016/j.joms.2006.10.005
- [8] James J. Xia, Liza Shevchenko, Jaime Gateno, John F. Teichgraber, Terry D. Taylor, Robert E. Lasky, Jeryl D. English, Chung H. Kau, and Kathleen R. McGrory. 2011. Outcome Study of Computer-Aided Surgical Simulation in the Treatment of Patients with Craniomaxillofacial Deformities. *Journal of Oral and Maxillofacial Surgery* 69, 7 (July 2011), 2014–2024. DOI:https://doi.org/10.1016/j.joms.2011.02.018
- [9] Warit Powcharoen, Wei-fa Yang, Kar Yan Li, Wangyong Zhu, and Yu-xiong Su. 2019. Computer-Assisted versus Conventional Freehand Mandibular Reconstruction with Fibula Free Flap: A Systematic Review and Meta-Analysis. *Plastic & Reconstructive Surgery* 144, 6 (December 2019), 1417–1428. DOI:https://doi.org/10.1097/PRS.0000000000006261
- [10] Sophie Bartier, O. Mazzaschi, L. Benichou, and E. Sauvaget. 2021. Computer-assisted versus traditional technique in fibular free-flap mandibular reconstruction: A CT symmetry study. *European Annals of Otorhinolaryngology, Head and Neck Diseases* 138, 1 (January 2021), 23–27. DOI:https://doi.org/10.1016/j.anorl.2020.06.011
- [11] Léonard Bergeron, Michelle Bonapace-Potvin, and François Bergeron. 2021. In-house 3D Model Printing for Acute Cranio-maxillo-facial Trauma Surgery: Process, Time, and Costs. *Plastic and Reconstructive Surgery - Global Open* 9, 9 (September 2021), e3804. DOI:https://doi.org/10.1097/GOX.0000000000003804
- [12] Benjamin Palla and Nicholas Callahan. 2021. Does the Use of Computer-Assisted Surgery Affect the Margin Status in Resections of Ameloblastoma? *Journal of Oral and Maxillofacial Surgery* 79, 7 (July 2021), 1467–1473. DOI:https://doi.org/10.1016/j.joms.2020.09.007
- [13] Jaime Gateno, James J. Xia, John F. Teichgraber, Andrew M. Christensen, Jeremy J. Lemoine, Michael A.K. Liebschner, Michael J. Gliddon, and Michaelanne E. Briggs. 2007. Clinical Feasibility of Computer-Aided Surgical Simulation (CASS) in the Treatment of Complex Cranio-Maxillofacial Deformities. *Journal of Oral and Maxillofacial Surgery* 65, 4 (April 2007), 728–734. DOI:https://doi.org/10.1016/j.joms.2006.04.001
- [14] Harry Schwartz. 2014. Does computer-aided surgical simulation improve efficiency in bimaxillary orthognathic surgery? *International Journal of Oral and Maxillofacial Surgery* 43, 5 (May 2014), 572–576. DOI:https://doi.org/10.1016/j.ijom.2013.10.018

- [15] Thiago de Santana Santos, Keylla Marinho Albuquerque, Marconi Eduardo Sousa Maciel Santos, and José Rodrigues Laureano Filho. 2012. Survey on Complications of Orthognathic Surgery Among Oral and Maxillofacial Surgeons. *Journal of Craniofacial Surgery* 23, 5 (September 2012), e423–e430. DOI: <https://doi.org/10.1097/SCS.0b013e31825e49c1>
- [16] Kari Panula, Kaj Finne, and Kyösti Oikarinen. 2001. Incidence of complications and problems related to orthognathic surgery: A review of 655 patients. *Journal of Oral and Maxillofacial Surgery* 59, 10 (October 2001), 1128–1136. DOI: <https://doi.org/10.1053/joms.2001.26704>
- [17] D.C. Bowe, E.A. Gruber, and N.M.H. McLeod. 2016. Nerve injury associated with orthognathic surgery. Part 1: UK practice and motor nerve injuries. *British Journal of Oral and Maxillofacial Surgery* 54, 4 (May 2016), 362–365. DOI: <https://doi.org/10.1016/j.bjoms.2016.01.026>
- [18] N.M.H. McLeod and D.C. Bowe. 2016. Nerve injury associated with orthognathic surgery. Part 2: inferior alveolar nerve. *British Journal of Oral and Maxillofacial Surgery* 54, 4 (May 2016), 366–371. DOI: <https://doi.org/10.1016/j.bjoms.2016.01.027>
- [19] N.M.H. McLeod and D.C. Bowe. 2016. Nerve injury associated with orthognathic surgery. Part 3: lingual, infraorbital, and optic nerves. *British Journal of Oral and Maxillofacial Surgery* 54, 4 (May 2016), 372–375. DOI: <https://doi.org/10.1016/j.bjoms.2016.01.028>
- [20] Joël Ferri, Charles Druelle, Matthias Schlund, Nicolas Bricout, and Romain Nicot. 2019. Complications in orthognathic surgery: A retrospective study of 5025 cases. *International Orthodontics* 17, 4 (December 2019), 789–798. DOI: <https://doi.org/10.1016/j.ortho.2019.08.016>
- [21] Hugo C. Gómez-Tone, Jorge Martin-Gutierrez, John Bustamante-Escapa, and Paola Bustamante-Escapa. 2021. Spatial Skills and Perceptions of Space: Representing 2D Drawings as 3D Drawings inside Immersive Virtual Reality. *Applied Sciences* 11, 4 (February 2021), 1475. DOI: <https://doi.org/10.3390/app11041475>
- [22] Ivan E. Sutherland. 1968. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I on - AFIPS '68 (Fall, part I)*, ACM Press, San Francisco, California, 757. DOI: <https://doi.org/10.1145/1476589.1476686>
- [23] Mohd Javaid and Abid Haleem. 2020. Virtual reality applications toward medical field. *Clinical Epidemiology and Global Health* 8, 2 (June 2020), 600–605. DOI: <https://doi.org/10.1016/j.cegh.2019.12.010>
- [24] Ashraf Ayoub and Yeshwanth Pulijala. 2019. The application of virtual reality and augmented reality in Oral & Maxillofacial Surgery. *BMC Oral Health* 19, 1 (December 2019), 238. DOI: <https://doi.org/10.1186/s12903-019-0937-8>
- [25] Ye-Joon Jo, Jun-Seok Choi, Jin Kim, Hyo-Joon Kim, and Seong-Yong Moon. 2021. Virtual Reality (VR) Simulation and Augmented Reality (AR) Navigation in Orthognathic Surgery: A Case Report. *Applied Sciences* 11, 12 (June 2021), 5673. DOI: <https://doi.org/10.3390/app11125673>
- [26] Hanna-Riikka Rantamaa, Jari Kangas, Sriram Kishore Kumar, Helena Mehtonen, Jorma Järnstedt, and Roope Raisamo. 2023. Comparison of a VR Stylus with a Controller, Hand Tracking, and a Mouse for Object Manipulation and Medical Marking Tasks in Virtual Reality. *Applied Sciences* 13, 4 (February 2023), 2251. DOI: <https://doi.org/10.3390/app13042251>
- [27] Federico Manuri, Federico Decataldo, Andrea Sanna and Paolo Brizzi. 2023. A Comparison of Two Interaction Paradigms for Training Low-Cost Automation Assembly in Virtual Environments. *Information* 14, 6 (June 2023), 340. DOI: <https://doi.org/10.3390/info14060340>
- [28] Duc-Minh Pham and Wolfgang Stuerzlinger. 2019. Is the Pen Mightier than the Controller? A Comparison of Input Devices for Selection in Virtual and Augmented Reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19)*. ACM, New York, NY, USA, Article 35, 1–11. DOI: <https://doi.org/10.1145/3359996.3364264>
- [29] Oğuzcan Ergün, Şahin Akın, İpek Gürsel Dino and Elif Surer. 2019. Architectural Design in Virtual Reality and Mixed Reality Environments: A Comparative Analysis. *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, Osaka, Japan, 914–915. DOI: <https://doi.org/10.1109/VR.2019.8798180>
- [30] Jari Kangas, Sriram Kishore Kumar, Helena Mehtonen, Jorma Järnstedt, and Roope Raisamo. 2022. Trade-Off between Task Accuracy, Task Completion Time and Naturalness for Direct Object Manipulation in Virtual Reality. *MTI* 6, 1 (January 2022), 6. DOI: <https://doi.org/10.3390/mti6010006>
- [31] Pat Dugard. 2014. Randomization tests: A new gold standard? *Journal of Contextual Behavioral Science* 3, 1 (January 2014), 65–68. DOI: <https://doi.org/10.1016/j.jcbs.2013.10.001>
- [32] Eugene Edgington, Eugene Edgington, and Patrick Onghena. 2007. *Randomization Tests* (0 ed.). Chapman and Hall/CRC. DOI: <https://doi.org/10.1201/9781420011814>
- [33] David C. Howell. 2013. *Statistical methods for psychology* (8. ed ed.). Wadsworth, Cengage Learning, Belmont, Calif.
- [34] Thomas E. Nichols and Andrew P. Holmes. 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. *Hum. Brain Mapp.* 15, 1 (January 2002), 1–25. DOI: <https://doi.org/10.1002/hbm.1058>
- [35] Flavio Wellington da Silva Ferraz, Liogi Iwaki-Filho, Gustavo Nascimento de Souza-Pinto, Lilian Cristina Vessoni Iwaki, An Tien Li, and Mauricio de Almeida Cardoso. 2021. A comparative study of the accuracy between two computer-aided surgical simulation methods in virtual surgical planning. *Journal of Cranio-Maxillofacial Surgery* 49, 2 (February 2021), 84–92. DOI: <https://doi.org/10.1016/j.jcms.2020.12.002>
- [36] Fabio Ritto, A.R.M. Schmitt, Thaís Pimentel, J.V. Canellas, and P.J. Medeiros. 2018. Comparison of the accuracy of maxillary position between conventional model surgery and virtual surgical planning. *International Journal of Oral and Maxillofacial Surgery* 47, 2 (February 2018), 160–166. DOI: <https://doi.org/10.1016/j.ijom.2017.08.012>
- [37] Sam Sheng-Pin Hsu, Jaime Gateno, R. Bryan Bell, David L. Hirsch, Michael R. Markiewicz, John F. Teichgraber, Xiaobo Zhou, and James J. Xia. 2013. Accuracy of a Computer-Aided Surgical Simulation Protocol for Orthognathic Surgery: A Prospective Multicenter Study. *Journal of Oral and Maxillofacial Surgery* 71, 1 (January 2013), 128–142. DOI: <https://doi.org/10.1016/j.joms.2012.03.027>