Touchless interaction with medical images based on 3D hand cursors supported by single-foot input: a case study in dentistry

Soraia Figueiredo Paulo[a,b], Filipe Relvas[a,b], Hugo Nicolau[a,b], Yosra Rekik[c], Vanessa Machado[d], João Botelho[d], José João Mendes[d], Laurent Grisoni[e], Joaquim Jorge[a,b], Daniel Simões Lopes[a,b,*]

[a] *INESC-ID Lisboa, IST Taguspark, Avenida Professor Cavaco Silva, 2744-016 Porto Salvo, Portugal*
[b] *Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal*
[c] *Université Polytechnique Hauts-de-France, CNRS UMR 8201, LAMIH, France*
[d] *Clinical Research Unit, CiiEM, Instituto Universitário Egas Moniz, Portugal*
[e] *Université de Lille, CNRS UMR 9189, CRIStAL, France*

Abstract

Feet input can support mid-air hand gestures for touchless medical image manipulation blueto prevent unintended activations, especially in sterile contexts. However, foot interaction has yet to be investigated in dental settings. In this paper, we conducted a mixed methods research study with medical dentistry professionals. To this end, we developed a touchless medical image system in either sitting or standing configurations. Clinicians could use both hands as 3D cursors and a minimalist single-foot gesture vocabulary to activate manipulations. First, we performed a qualitative evaluation with 18 medical dentists to assess the utility and usability of our system. Second, we used quantitative methods to compare pedal foot-supported hand interaction and hands-only conditions next to 22 medical dentists. We expand on previous work by characterizing a range of potential limitations of foot-supported touchless 3D interaction in the dental domain. Our findings suggest that clinicians are open to use their foot for simple, fast and easy access to image data during surgical procedures, such as dental implant placement. Furthermore, 3D hand cursors, supported by foot gestures for activation events, were considered useful and easy to employ for medical image manipulation. Even though most clinicians preferred hands-only manipulation for pragmatic purposes, feet-supported interaction was found to provide more precise control and, most importantly, to decrease the number of unintended activations during manipulation. Finally, we provide design considerations for future work exploring foot-supported touchless interfaces for sterile settings in Dental Medicine, regarding: interaction design, foot input devices, the learning process and camera occlusions.

Keywords:
Foot, Hand, Gestures, Asepsis, Dentistry, User-Computer Interface

1. Introduction

Given the expanding usage of touchless technology in sterile settings such as the operating room (OR), it has become critical to ponder how gesture-controlled medical image interaction can be designed to support clinicians. In Dental Medicine, medical dentists rely on image data to confirm their diagnosis and perform surgical procedures, such as dental implant placements or tooth extractions [1], which often resort on foot switches for assisted control of surgical equipment. Most commonly, medical dentists adopt a sitting working posture, but depending on the demands of the procedure, they can be required to perform it while standing [2]. Dental surgery procedures occur in a sterile setting, where patients sit or lie on a dental chair, surrounded by the components of the dental unit, namely display monitors with 2D and 3D medical images [1, 3]. While 2D images can be visualized in three different planes (axial, coronal an sagittal) to evaluate bone volume and quality, 3D models offer the clinician an overall perspective of the anatomical structures [4].

*Corresponding author
Email address: daniel.lopes@inesc-id.pt (Daniel Simões Lopes)

To have direct control over 2D and 3D images, the medical dentist may need to move away from the patient, interacting physical input devices covered with a surgical cloth, or changing gloves to maintain asepsis, ultimately delaying the procedure [3]. Although previous work has proposed touchless image manipulation techniques [5, 1, 3, 6, 7, 8], little attention has been given to using such interfaces in the dental domain, much less to 3D touchless interfaces leveraged by minimalist feet input in dental settings. Indeed, the lack of guidelines to assist the design of such interfaces is a problem that still needs to be addressed in order to promote more effective practices and reduce dental treatment duration [7]. The problem with touchless manual input is that it can be imprecise and error-prone due to unintended activations especially at the end of a manipulation.

In this paper, we investigate using single-foot input to support 3D hand cursors for touchless medical image interaction in dental settings. We aim to answer key research questions such as: what are the benefits and drawbacks of foot-supported 3D touchless image manipulation? How can these technologies be included in the daily clinical practices of medical dentists? Can foot device positioning positively affect medical image access? Besides 2D data, is consulting with 3D data also relevant for the medical dentist? To address these questions, we first built a foot-supported 3D gesture-controlled image manipulation system. The prototype exploits minimalist single-foot interaction in both standing and sitting positions, which correspond to the body postures adopted by medical dentists in sterile clinical settings. We then conducted a qualitative user-study with semi-structured interviews, using the prototype as a design probe to assess the professionals' experiences and expectations. This was followed by a quantitative study, in order to investigate how foot-support affects the performance of touchless 3D gesture interaction. Our main contributions are (i) a prototype meant to explore a design space that extends beyond 2D manipulation by enabling 3D interaction; (ii) a professional assessment with medical dentists; and (iii) design guidelines for future work on foot-supported touchless interaction in dental settings.

2. Related Work

Over the last decades, touchless interfaces have been increasingly adopted in clinical settings [5, 7, 8], while at the same time they provide interesting interaction techniques for several distant viewing and content manipulation applications [9, 10, 11, 12]. These technologies open novel opportunities for surgical applications, where prompt access to anatomical imagery is key for a successful procedure in an environment where sterility is mandatory [13, 7, 14].

Within sterile clinical settings, gesture-driven approaches have been widely explored to produce image manipulation and navigation interfaces. The use of depth cameras, such as Microsoft's Kinect, has been applied to 2D [15] and 3D medical image interaction [3, 6, 16], improving task completion time and spatial awareness. Other approaches have resorted to wearable RGB-D sensors to enable touchless interaction [17, 18], as well as Leap Motion's infrared stereo camera [1]. This included emulating the use of mouse and keyboard [17], or enabling 3D manipulation for preoperative planning and surgical navigation [1, 18]. However, none of these works provides tangible guidelines resulting from the assessment of foot-supported 3D hand interaction by medical dentists, especially targeting minimalist feet vocabulary.

Depending on the complexity of 2D and 3D interaction tasks, specific input modalities, such as voice, gaze or foot control, can perform differently in sterile clinical settings [19]. Recently, gaze has been combined with foot input [20] and even auditory feedback [21]. Still, gaze and foot input were found to easily interfere with each other [20], while auditory feedback was highly impacted by the range of sound sources inside an OR [21]. Feet input, however, was found to be suitable for low accuracy and quick spatial tasks [22], along with soft tasks in hands-busy situations [23]. While simple foot tapping on foot pedals enables fast on-screen content selection [24], the use of more subtle gestures, namely single-foot heel rotations [25] and sequential foot tapping [26, 27], allows the user to stand in a stable posture. Hence, this suggests that feet would be more appropriate as a complementary gestural input method, namely to provide control over activation events [28]. In fact, many researchers have explored feet as a medium to support hand gestures for selecting modes and controlling a camera in a 3D modelling application [29], enabling the assembly of virtual 3D objects [28], selecting menu options [30], and three-dimensional navigation, selection, manipulation and system control tasks using a depth sensing camera [31].

Closer to our work, was the use of foot-supported hand interaction to manipulate 2D medical images in surgical settings [32]. Although we share similar interaction contexts and investigate single-foot gestures for activation events, namely screen selection, the previous work depends on hand wearables and dedicated hardware for foot gesture recognition. Also, exploring the foot to select and switch screens required a more choreographic interaction, whereas we privilege minimalist foot vocabulary only to confirm hand selection and enable interaction. Furthermore, their main content are 2D images, while we include both 2D and 3D image data. Thus far, the potential impact of non-wearable, foot-supported 3D touchless interaction for sterile image manipulation has yet to be investigated in dentistry settings.

To the best of our knowledge, only [1, 3] focused the dental domain, producing insights into this particular setting which argue in favor of touchless interfaces. The pilot study by [1] reported the use of a gesture-based 2D and 3D image interaction system during dental surgery procedures. The prototype was integrated with the dental unit chair and recognized touch-like gestures on a virtual vertical surface, allowing the medical dentist to remain next to the patient while interacting with image data. On the other hand, [3] considered two-hand gestures performed while the user was standing. In order to manipulate 3D models, the system included 7 unique gestures, which lead to increased difficulties to learn the gesture vocabulary. Yet, none of these works accounted for unintended actions, nor did produce guidelines for 3D hand interaction.

Although we considered hand gestures for 2D and 3D image browsing and manipulation, along with feet gestures for activation events, there still continues to exist a literature gap as the vast majority of studies investigate 2D cursor based interfaces and rely on more choreographic feet gesture vocabulary [27, 32, 20]. We build upon previous work [6], in order to address the limitations imposed by unintended activations, affecting the interaction's precision. Our rationale for hand and foot interaction sought simple gestures, easy to track and to remember [9, 5, 8]. Following the works of [33, 22] and considering the design guidelines of [25, 34] that explored the interaction potential of single-foot input, our prototype draws on the strengths of appropriate typologies of foot gestures [33] to support mid-air hand gestures while manipulating virtual objects, without depending on expensive dedicated hardware. Our study adds to the state of the art by addressing the potential of using hands as true 3D cursors supported by minimalist foot gesture input for both standing and sitting positions. Thus, we investigate the potential effects of using foot-supported touchless 3D manipulation techniques on dentistry settings in order to produce design guidelines fitting to this scenario.

3. Touchless interaction with 3D hand cursors and single-foot support

To understand the potential of touchless interaction based on 3D hand cursors supported by minimalist single-foot input, we built *TOOTHFAIRY* (Touchless interaction with single-fOOT support of 3D Hand cursors For Asepsis In dentistRY). This prototype was designed to browse and manipulate 2D and 3D medical images in mid-air through 3D hand gestures, introducing more precise control over activation events using a small set of simple gestures. Gesture recognition is carried out using a depth camera to capture the position of the hands and detect hand gestures, at the same time foot input is either provided via an optical marker detected by the camera's infrared sensor or through a foot pedal, while standing and sitting, respectively (Figure 1).

3.1. Graphical User Interface

The graphical user interface consists of a 2x2 layout of four distinct viewports (Figure 2), each corresponding to a different projection of the object to manipulate: three orthographic projections (axial, sagittal and coronal), along with a perspective projection (3D image). The user may also choose to maximize one of the viewports, as the interface becomes a 1x1 layout displaying the selected projection. The 3D view in particular, relies on a 3D grid to enhance depth perception, as well as to create a sense of relative dimension. Each viewport is limited by a colored window frame, which becomes brighter to indicate that viewport was selected. In addition, the hands' positions are mapped to the display and represented by two white hand knobs.

3.2. 2D and 3D Image Manipulation

In order to enable image manipulation, users must place the cursor representing the dominant hand on the viewport they wish to interact with, and confirm this selection using specific gestures. The set of features made available to control the

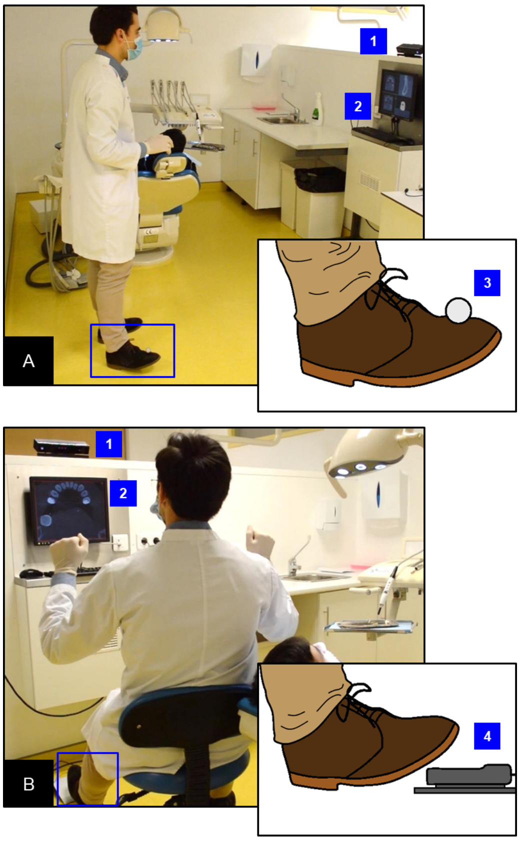

Figure 1: The setup of the *TOOTHFAIRY* prototype includes (1) a depth camera to capture gesture interaction and (2) a display monitor for the graphical user interface. Regarding feet input, the prototype considers (A) a standing condition, which requires the use of (3) an optical marker, and (B) a sitting condition, which uses a foot pedal instead.

interface depends on whether the viewport users selected corresponds to a 2D or 3D view. In 2D views, users have access to unconstrained translation, scale, and 2D slice navigation, whereas the 3D view also enables rotation and offers the ability to perform constrained transformations (i.e., transformations along an axis). To perform a constrained transformation, users are required to use both hands to define the axis around which the transformation will occur first (Figure 3), and then perform the gesture corresponding to the transformation they wish to apply.

3.3. Volume data

A single dental Cone Beam Computed Tomography (CBCT) anonymized dataset provided by our clinical partners was used. Image dimensions correspond to 512x512 pixels in a volume of 512 slices.

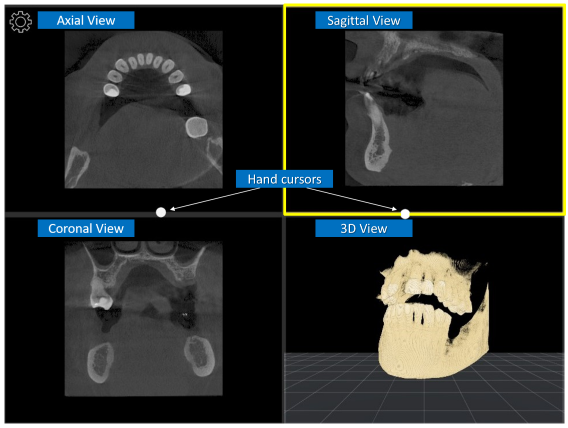

Figure 2: The *TOOTHFAIRY* interface is composed of two white hand knobs and four viewports: Axial (top-left), Sagittal (top-right) and Coronal (bottom-left) Views; 3D Volume (bottom-right). The colored window frame indicates which viewport is selected (in this case, the sagittal viewport).

3D images were generated by a built-in volume rendering engine [35], which reconstructs 3D data from the stacks of 2D CBCT slices and renders it using a raymarching shader.

4. Methodology

To evaluate the potential impact of using touchless 3D hand cursors supported by single-foot input, we followed an iterative methodology. First we performed a qualitative assessment and investigated two different feet input conditions using the *TOOTHFAIRY* prototype. Then, based on our findings (see subsection 5.1), we developed *TOOTHFAIRY 2.0* and performed a quantitative assessment of hands-only versus feet-supported hand interaction.

4.1. Qualitative Assessment

We conducted an interview study with 18 medical dentists using *TOOTHFAIRY* as a design probe. The prototype was set up in a meeting room used for discussing clinical cases, where there were no dental chairs nor surgical equipment.

Based on our research questions, we considered that: (i) Consulting with 3D data is relevant for the medical dentists; (ii) Touchless 3D (and 2D) image manipulation with hand gestures benefits from single-foot input; (iii) 3D cursors are useful for manipulating 3D medical data; (iv) *TOOTHFAIRY* is beneficial for the dental domain.

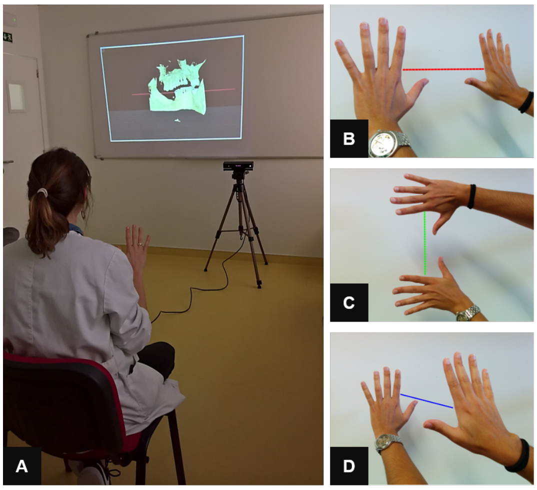

Figure 3: Constrained Transformations: (A) 3D image manipulation around the x-axis. Hand-gestures for axis selection around (B) the x-axis; (C) the y-axis; (D) the z-axis.

4.1.1. Implementation

While we considered two feet input conditions, hand interaction remained unaltered for both standing and sitting positions. Three possible hand-gesture states are available: opened, closed and lasso. If one of the hands is closed, the user can either translate the selected object, or navigate the 2D slices, depending on the movement's direction towards the screen: parallel or perpendicular, respectively. In case both hands are closed, once they move away or towards each other, the image's scale increases or decreases, accordingly. The lasso hand state is used for rotating the 3D image, as the neutral state is represented by open hands with their palms facing the screen. Hand-gesture manipulation is illustrated in Figure 4(a).

Our feet gesture vocabulary is designed around a minimalist set of gestures performed with the dominant foot, while the non-dominant foot lies flat on the floor, remaining static to ensure proper weight balance. We designed two different feet input conditions given two scenarios frequently found inside dental settings: (i) the medical dentist performs the procedure standing; or (ii) sitting closely to a patient on a dental chair. While standing, using an optical marker on the foot, allowing users to be mobile, we consider their dominant foot tapping and swivel rotation to perform window selection (angle of rotation between 0^{o} - 45^{o}) and window resizing (45^{o} - 90^{o}). While sitting, this corresponds to a click on the foot pedal to perform window selection, and double-click for resizing. To stop interaction, the medical dentist must perform the same gesture used for window selection. Feet gestures are illustrated in Figure 4(b).

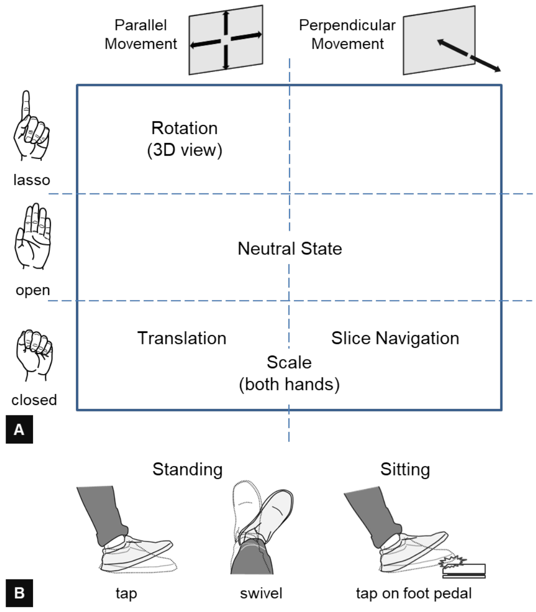

Figure 4: Interaction Space: (a) User actions mapped according to hand gesture and movement's direction relative to the screen of the proposed touchless interaction system; (b) Feet gestures considered while the user is standing (tap and swivel) and sitting (tap on the foot pedal)

Regarding our mapping ratios, we consider: i) relative mapping for hand translation and hand rotation, 3D hand position is projected to screen coordinates and distance covered between frames is mapped to move or rotate object; ii) absolute mapping for feet rotation and relative mapping for foot translation (1^{o} corresponds to 1 cm); iii) the relative position of the marker was used to detect tapping, as swivel rotation was calculated using the forward vector perpendicular to the user's chest and the foot vector; iv) mapping functions are linear, otherwise users would lose task precision, forcing them to learn and adjust their gestures to a non-linear mapping. The design rationale behind the hand gestures is that they need to be simple, easy to track and remember.

4.1.2. Participants

Eighteen medical professionals (1 maxillofacial surgeon and 17 medical dentists), took part in our study (6 female and 12 male). Their specialized experience ranged from 1 to 15 years and they always

(61%) or regularly (39%) use radiographic images to perform surgical planning, although 3D applications are mostly only occasionally (67%) used. Six participants reported previous experience with spatial gesture interaction devices, such as Nintendo's Wii Remote, Playstation Move, and Microsoft Kinect.

4.1.3. Apparatus

Our setup used the skeleton provided by the System Development Kit (SDK) of Microsoft Kinect V2 depth camera to detect the hands' positions. Given the standing and sitting scenarios, we considered two different feet input conditions: (i) using an optical marker placed on the dominant foot, to allow the depth camera to detect both height variation, with respect to the ground level position (i.e. foot tapping), and relative angular heel rotation (i.e. foot swivel); (ii) using a stapler connected to a Makey Makey V1.2, considered as the pedal that served as a switch to enable single-foot input (Figure 5). The depth camera and the Makey Makey were both connected to the same laptop computer where TOOTHFAIRY was running (Asus ROG G752VS, Intel® Core™ i7-6820HK Processor, 64GB RAM, NVIDIA GeForce GTX1070). All the code was developed in C# using Unity game engine (version 2017.3.0f3).

4.1.4. Procedure

Participants were prompted to browse through the images and observe the case, exploring all interaction features while describing what they were seeing and experiencing, considering the potential use of the system in a sterile clinical setting. This was done both in standing and sitting positions. Conditions were counterbalanced, as half the participants started in the standing condition and the other half sitting. Each session lasted approximately 30 minutes. Before starting, users were asked to fill a demographic questionnaire and were introduced to the experiment regarding the features and the scenarios we wanted to explore.

At the beginning, participants were asked to select a viewport, which they could either choose to use as a default sized window or maximized. From there, they were asked to proceed as if they were in an image-dependent intraoperative setting, where image browsing and manipulation tasks are essential for medical dentists. Once all interaction features were tested and participants were finished giving feedback about the first condition they were

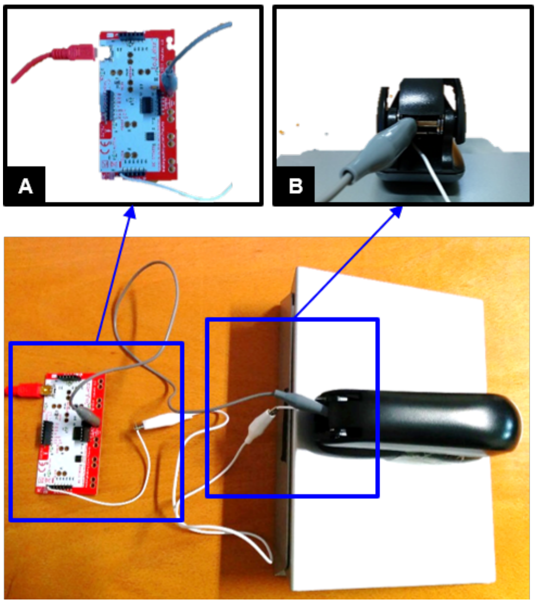

Figure 5: TOOTHFAIRY setup using Makey Makey: (A) the Makey Makey board is connected via a USB cable (red) to the laptop PC; the white cable is connected to the board's keyboard input, while the grey cable is connected to the board's ground; (B) each cable is attached to a metallic part of the stapler, one to the upper (grey) and one the lower part (white), to allow the stapler to work as a switch.

testing, they were asked to fill in a questionnaire. The goal was to rate each feature in terms of usefulness (Is this useful?), execution (Is this performed in an appropriate manner?), memorability (Is it easy to remember how it is performed?) and usability (Is it easy to achieve the desired result?). The same procedure was then repeated for the second condition. Participants were encouraged to think-aloud while using the prototype. Finally, we conducted a semi-structured interview in order to obtain additional feedback, mainly concerning user preferences and prototype's potential and/or viability.

4.1.5. Measures

Throughout the session, one researcher gathered observational notes regarding user's experiences and expectations. The analysis of the data collected during the study was open coded in a process which resulted in 21 codes. To obtain an overall opinion of user's preferences and satisfaction, we conducted a 6-point Likert-scale questionnaire, where 1 meant the user totally disagreed with the statement, and 6 the user totally agreed with it. A Wilcoxon signed-rank test was conducted to compare the standing and sitting conditions.

4.2. Quantitative Assessment

Considering the results from the qualitative study (see subsection 5.1), this experiment aimed to compare the performance of *TOOTHFAIRY 2.0*, using hand input supported by a foot pedal, and a hands-only interaction technique. The latter served as a baseline condition, exploring the same set of features and gestures described for our first prototype, while using the non-dominant hand to control window selection. To this end, we used the same apparatus described in the qualitative study (see subsubsection 4.1.3), switching the stapler for a standard foot pedal, also connected to the Makey Makey board. The prototype was set up in a dental treatment room (Figure 6), using the monitor display available in that space.

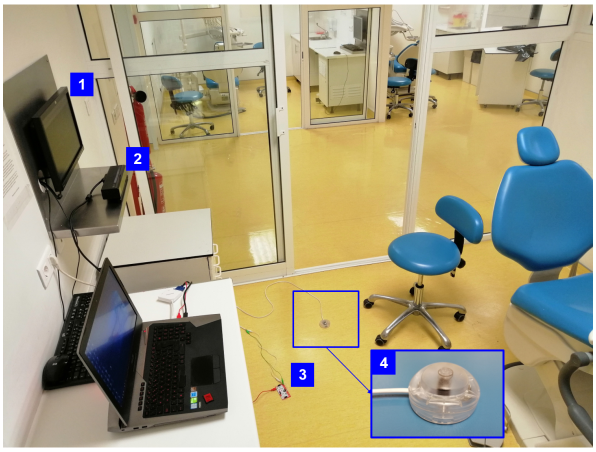

Figure 6: TOOTHFAIRY 2.0 setup in the dental treatment room: 1) Monitor display; 2) Depth camera; 3) Makey Makey board; 4) Foot pedal.

4.2.1. Implementation

In *TOOTHFAIRY 2.0*, we consider a single-foot input condition, using a foot pedal in a sitting position. In this case, the dominant foot is further explored to support activation events, providing more effective control input [22], in order to avoid undesired position/angular displacements whenever users perform hand releases.

This prototype includes the same set of features described for the first prototype, but with a significant difference: the dominant foot controls mode switching between transformations. To this end, we consider three feet gestures: long click (i.e. click and hold for 2 s), click and double click. In order to resize a viewport, users must place their dominant hand on the viewport and perform a long click. Image manipulation is only available in maximized viewports. To enable any transformation, at least one hand must be on or above the line defined by the shoulder joints. In this case, if the user performs a foot click and moves the hand parallel to the screen, translation is activated. However, if the hand is moved perpendicularly towards the screen, slice navigation is activated instead. If both hands are on or above the shoulder line, the user can scale the volume by performing one click and moving the hands towards or away from each other. Also, to rotate the 3D image, users must perform a double click. Constrained manipulation was also available. To stop manipulation, users must always perform the same foot gesture used to start (i.e. click to stop translation/slice navigation/scale, double click to stop rotation).

4.2.2. Participants

Twenty two Medical Dentistry professionals (13 female, 9 male) participated in our quantitative study. These included one medical dentist with 1 year of specialized experience in Orthodontics, and 21 Medical Dentistry finalist students. Most participants always (63.6%) or regularly (31.8%) use radiographic images for surgical planning, while 3D applications are mostly occasionally (50%) or never (36.4%) used. Finally, five participants reported previous experience with spatial gesture interaction devices.

4.2.3. Procedure

First, researchers introduced the project and outlined the goals of the session. Participants were asked to fill out a demographic questionnaire and an informed consent prior to starting. Every participant performed 6 tasks both in a hands-only condition, which was considered the baseline condition, and in the TOOTHFAIRY condition, while sitting in front of a display. Conditions were counterbalanced, as half the participants started with the hands only condition and the other half with the TOOTHFAIRY condition. Tasks were performed in a randomized order to mitigate learning effects. All sessions followed the same structure and lasted approximately 30 minutes. In each condition, participants were shown how to use the prototype and had a training period of a maximum of 8 minutes, during which they could ask questions while exploring the prototype. During this period, users were asked to perform two training tasks, one in a 2D viewport and one in a 3D viewport. Then, users were prompted to perform a set of 6 tasks, as each individual task had to be completed within a maximum of 2 minutes. If the time limit was reached, we considered the attained position, orientation and scale as final. Although some tasks required only translation or rotation transformations, none of these were restricted on any task. Since the separation of degrees-of-freedom has been shown to benefit precision in spatial manipulations [36], users were asked to use translation and rotation in their constrained form (along and around the axis, respectively).

4.2.4. Tasks

Participants were asked to perform a set of 6 tasks per condition, which represent potential imagery manipulations in Medical Dentistry clinical practice. There were 3 tasks regarding 2D content (tasks A to C), as well as 3 tasks concerning 3D content (tasks D to F) with different levels of complexity, according to the number of transformations required. In task A the goal was to select a designated slice in the axial viewport. In task B, the objective was to place the current 2D image on the 2D square target, in the sagittal viewport, by translating it along the X axis and scaling it. Task C required the user to select a given slice, in the coronal viewport, and place it on the target by translating it along the Y axis and scaling it. In tasks D to F, the goal was to place the 3D model's mandible on the 3D target mandible. Task D required the user to rotate the 3D model around the Y axis. Task E required the model to rotate around the X axis and translated along the Y axis. Finally, task F needed the model to rotate around the X axis, translated along the X axis and scaled. Each task was finished by resizing (i.e. minimizing) the active viewport or as the result attained by the end of the time limit. At the end of both tasks, users were asked about which condition they preferred to use.

4.2.5. Measures

All data was recorded in a log file for further analysis. For each participant and task, we computed time to complete task (i.e. time elapsed between the moment a viewport was maximized and the end of the last transformation), distance to target position and rotation (i.e. difference between the position/rotation of the target and the 3D mandible manipulated by the user), and scale factor error (the scale factor value consists in the ratio between the object's initial scale and its current scale. Thus, the scale factor error is the absolute difference between the target scale factor and the 3D mandible's final scale factor). In addition, we registered the number of unintended activations (i.e. accidental window resizing, unintended object displacement/rotation during its release, incorrect gesture detections). Since data did not follow a normal distribution (Shapiro-Wilk test, $p < .05$), we used the nonparametric Wilcoxon signed-rank test to compare *TOOTHFAIRY 2.0* and baseline conditions.

5. Results

In this section we present the results of our qualitative and quantitative assessments.

5.1. Qualitative Assessment

For the complete set of results, please see Table 1 in Appendix. Statistically significant results are presented in Figure 7.

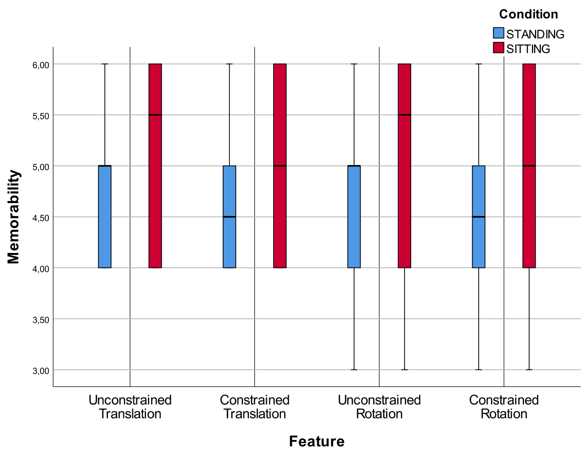

Figure 7: Questionnaire results regarding Memorability, comparing the use of the optical marker and the foot pedal for Unconstrained Translation, Constrained Translation, Unconstrained Rotation and Constrained Rotation features.

In general, all image manipulation tools were considered useful in both conditions. Users agreed that all features had an adequate form of being executed, also receiving similar classifications in terms of usability. Results indicate there were no significant differences between constrained and unconstrained manipulation.

Regarding foot interaction, the lack of statistically significant differences between the two input conditions in most criteria suggested there was not a clear preference for either approach. However, users considered it was easier to remember how to perform unconstrained translation ($p = 0.034$), constrained translation ($p = 0.034$), unconstrained rotation ($p = 0.020$), and constrained rotation ($p = 0.020$), while sitting than while standing, since using the pedal only required one or two clicks, instead of a combination of foot tapping and swivel.

Our findings from the design probe complement and expand on the questionnaires' information. From our analysis, five themes emerged from the interview study: (i) *Learning effort*; (ii) *From current technologies to 3D features*; (iii) *Familiar devices are preferred, but new opportunities emerge*; (iv) *Environmental barriers*; and (v) *Novel surgical practices.*

Learning effort: "This will involve a learning curve, so that we are able to move more naturally and evaluate everything necessary."

The theme of the learning effort arose frequently. Most users expressed a concern with gesture learnability, as memorizing gestures and a *"choreography"* was seen as a potential limitation to such interfaces. Thus, this indicated a gesture-based approach would need to be simple and natural, in order to be more easily and quickly adopted. That was not the case for axis operations, as many users felt it was difficult to use, given the gestures required to define the manipulation axis. Something that was also pointed out was the cost of transitioning from a mouse to a gesture-based interaction, what would it involve and how smooth could it be, so that it would not add to the learning curve of the new interface. It was made clear that this should imply no additional effort, specially in an already demanding scenario such as the OR.

From current technologies to 3D features: "We are able to do everything we need."

This theme outlines how users see moving forward to a gesture-based approach and what they think this would add to existing technologies. While users were open to exploring new options, they often referred to current tools. As a result, it was often stated that the set of features that was introduced in the novel interface had to match existing functionalities. However, according to the particularities of each participant's specialty, it was apparent they were willing to leverage the possibilities offered by using the hands as 3D cursors, namely to manipulate 3D objects, as they suggested *"accessing the segmented model"* or *"using the 3D view to get a better perspective of the patient's bite, to improve the analysis of dental occlusions"*. This is also suggested by the positive responses regarding our 3D interaction design. Ideally, medical dentists believed this opportunity should build upon conventional systems in order to make it viable: *"At the clinic this would be viable if we could have a large display, working with CBCT* (Cone Beam Computed Tomography) *data, along with an easy access to the application"*, said a medical dentist regarding the system used at his dental practice.

Familiar devices are preferred, but new opportunities emerge: "The foot pedal comes in more handy, it is what we are used to."

The theme of familiarity was mentioned throughout the sessions, explaining how and why most medical dentists preferred using a foot pedal, rather than an optical marker. To begin with, users seemed to understand the purpose of a foot-supported interaction, as a means of relieving the cognitive load of controlling a touchless interface: *"A hands-only approach would be good, but maybe it would become too confusing. It is better to have the pedal, it makes the hand gestures simpler"*. The explicit preference for the pedal was often observed: *"It is what we are used to"*, mostly as if it was a requirement for viewport selection and resizing to be considered easy or comfortable to perform. Given the frequent use of foot devices in dental clinical practices, it becomes easier to adapt to adding a new functionality to the pedal, other than going for a new device, which may be seen as a more abrupt transition. This was also suggested by the memorability results of our questionnaire. In spite of choosing more familiar options, users also noticed that something like a foot marker could bring up new opportunities: *"Since the pedal represents having a physical device, using it becomes easier, but the nonexistence of pedals might be the future"*. The idea of moving more freely around the room was found to be the main advantage of our standing condition, leading three medical dentists to consider this a promising approach that enabled them to adopt different working positions during a procedure.

Environmental barriers: "A head surgery involves at least 3 to 4 people, so there would be camera occlusions."

This theme outlines the obstacles imposed by the current practices, social and physical environments where sterile image manipulation would be considered. ORs and other sterile clinical scenarios involve accounting for a number of healthcare professionals working in reduced physical environments. Therefore, users highlighted that the prototype interface would need to adjust to a setup that requires little physical space, in order to fit several sterile scenarios which are already filled with large-sized equipment, namely the dental chair. Also, the maxillofacial surgeon raised awareness of the fact that in a social environment with a minimum of 3 to 4 people moving, the system design should consider how to avoid that at any given moment someone could occlude the camera and lead to unintended actions. For example, one participant suggested implementing a Virtual Reality (VR) scenario, as a way of requiring less physical space to see and manipulate image data: *"In an intraoperative setting, the assistant could put the Head-Mounted Display (HMD) on the surgeon, he could see and manipulate the images and then the assistant could take it off"*. Another concern was the need to cover or isolate the foot devices, as they expressed their worries about what would happen if any blood or lavage fluids were spilled over the device.

Novel surgical practices: "This would be extremely useful for dental implant placement!"

In this theme we describe the potential applications of foot-supported touchless medical image interaction to novel surgical practices, according to the vision of our participants. As expected, one of the main advantages observed while using a touchless interface was its convenience. Direct manipulation of mouse-based input forces the user to change gloves while touchless manipulation avoids both this and possible contamination. In addition, medical dentists considered it would be a valuable tool during different intraoperative contexts, affirming that *"Orbital surgery would love this!"*. Many complex surgeries benefit from accessing 2D and 3D images, which enable the surgeon to get a better perception of the individual's anatomical variability and to identify specific areas or structures more clearly. The more natural and faster the access is, the better. In this line of view, users also suggested adapting the prototype to an educational setting, where the apprentice would be able to follow a procedure more easily by being provided additional information to help him understand the anatomy. The same applies to presenting or discussing clinical cases. Finally, medical dentists expanded on the idea of intraoperative image manipulation. In that case, they envisioned a simulation scenario, where the images would display a representation of the surgical instruments, enabling an image-guided procedure.

5.2. Quantitative Assessment

For the complete set of results, please see Table 2 to Table 7 in Appendix. Statistically significant results are presented in Figure 8, 9, 10, 11 and 12.

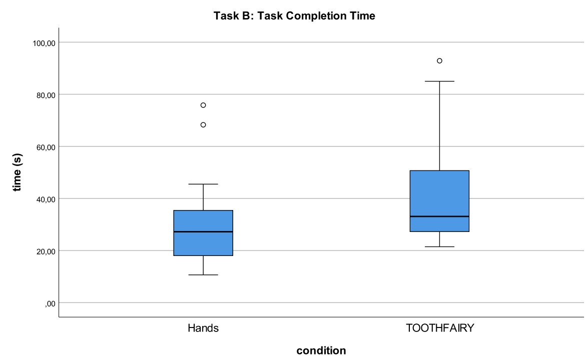

Figure 8: Task completion time for Baseline (hands-only) and TOOTHFAIRY conditions in Task B

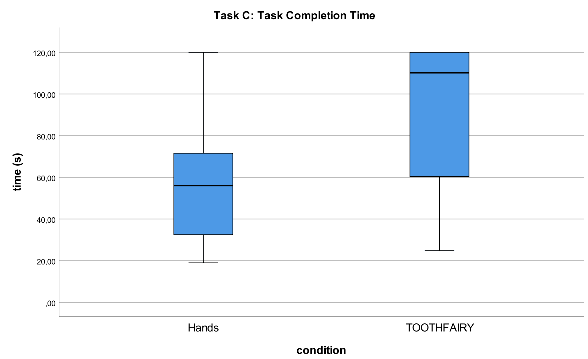

Figure 9: Task completion time for Baseline (hands-only) and TOOTHFAIRY conditions in Task C

In Task A (2D slice navigation), the scale factor error was significantly lower in the TOOTH-

FAIRY condition (Hands: Mean (M)=.24 , standard deviation (s.d.)=.44; TOOTHFAIRY: M=.00, s.d.=.00; $Z = -2.366$, $p = 0.018$). Considering Task A did not require any scale transformation, this suggests that in the baseline condition there were unintended activations leading to this result. Also, in Task B (2D translate-scale) and Task C (2D translate-scale-slice navigation) task completion time was significantly higher in the TOOTHFAIRY condition (Task B, Hands: M=30.20 s, s.d.=17.23 s; TOOTHFAIRY: M=41.83 s, s.d.= 20.57 s; $Z = -2.576, p = 0.010$; Task C, Hands: M=60.31 s, s.d.=28.98 s; TOOTHFAIRY: M=93.42 s, s.d.= 30.66 s; $Z = -2.678, p = 0.007$), which was indicative that the more controlled manipulation technique TOOTHFAIRY aims at, could also imply a slower process.

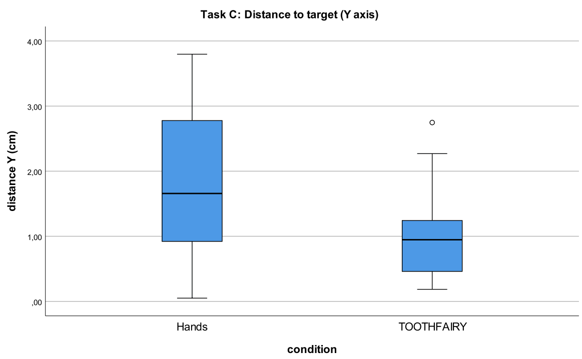

Figure 10: Distance-Y to the target for Baseline (hands-only) and TOOTHFAIRY conditions in Task C

In addition, Task C reported a significantly lower distance to the target in the Y axis (Hands: M=1.83 cm, s.d.=1.20 cm; TOOTHFAIRY: M=1.06 cm, s.d.= .73 cm; $Z = -2.415, p = 0.016$).

Finally, in Task D (3D rotate) and Task F (3D rotate-translate-scale), the rotation error in the Z axis was significantly lower in the TOOTHFAIRY condition (Task D, Hands: M=3.97°, s.d.=4.96°; TOOTHFAIRY: M=1.25°, s.d.= 2.09°; $Z = -1.988, p = 0.047$; Task F, Hands: M=1.89° s, s.d.=2.41°; TOOTHFAIRY: M=.66°, s.d.= .99°; $Z = -2.069, p = 0.039$). In the remaining tasks and measures, results indicate there were no statistically significant differences between both conditions.

Considering the ratio of unintended activations between both conditions, TOOTHFAIRY produced approximately 3.5 times less unintended activations than the hands-only condition. Even though most users acknowledged this when asked about their

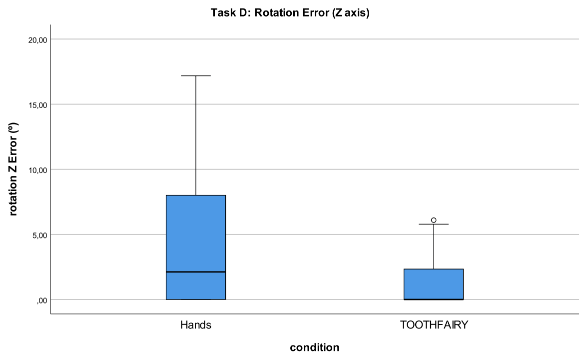

Figure 11: Rotation-Z Error for Baseline (hands-only) and TOOTHFAIRY conditions in Task D

Figure 12: Rotation-Z Error for Baseline (hands-only) and TOOTHFAIRY conditions in Task F

preference, 57.1% still preferred only using their hands, mainly because *"It is quicker"* than combining hands and feet for this purpose. On the other hand, users preferring foot-supported hand interaction stated that *"I am already used to combining hands and feet. Using a hands-only approach requires me to think more and keep these new gestures in mind"*, also reinforcing how it avoided a number of unintended actions that occurred in the other condition, *"When I wanted to stop* (manipulating the image), *sometimes I ruined everything I had done*(referring to the hands-only condition). *While using the pedal, I could keep my hands still for a moment, pressing the pedal, and after that I could move and everything was kept in place."* Nonetheless, users also mentioned that using TOOTHFAIRY's pedal could be easier if it was not as different from the pedals they are mostly used to, which should be kept on hold to remain active, instead of a start and stop-like button.

6. Discussion

We use our findings to discuss design guidelines for single foot-supported touchless image manipulation for the dental field based on clinicians' point of view. Our intention is to complement other sets of guidelines suggested by previous work in the literature by providing new insights in this application domain. The main issue that foot input aims to tackle in this study is that of unintended activations at the end of a command/manipulation. Indeed, using hands alone to specify both image manipulation and start/stop of actions leads to imprecise control and difficulties in specifying the end of an interaction. Often, this translates to muitple painstaking interactions to achieve the desired state.

6.1. Implications for interaction design and interfaces

In sterile clinical scenarios, such as the spaces where dental surgery procedures occur, touchless interfaces were found to avoid possible contamination, allied to reduced delays and miscommunications that arise when medical professionals need to see and manipulate images [5]. Besides being more convenient, clinicians saw touchless interaction as an opportunity to further extend the functionalities of current systems to touchless 3D interaction. This would not only enhance the aid provided by such images during complex surgeries, but also create an active teaching scenario. In TOOTHFAIRY, foot-support was regarded as a way to avoid unintended actions, at the cost of a more pragmatic and quicker alternative, such as a hands-only interaction technique. Even though users had less unintended activations with TOOTHFAIRY v2, results suggest that they also recovered quicker from errors while using a hands-only approach, than using the more discrete process proposed in TOOTHFAIRY, in the sense that transitions between transformations are well marked, contrary to what happens in the hands-only condition, where transitions are smoother. When designing for sterile clinical scenarios, we recommend: (i) Image manipulation should feature translation, rotation (3D images only), scale and slice navigation, such as [32, 15, 6]; (ii) Feet interaction should explore simple gestures to enable/disable manipulation, while the active window remains selected; (iii) Interaction design needs to consider metaphors that are domain-specific, allied to a balanced trade-off between smooth transitions and effective mode switching that avoids unintended actions; (iv) Interfaces would benefit from accessing the Picture Archiving and Communication System (PACS) used at the medical facility, in order to facilitate the access to medical imagery. Close to that is [17] that by transforming pointing gestures into general events, enables the surgeon to control numerous medical systems, including the PACS.

6.2. Implications for the choice of foot device

While some procedures may require the clinician to adjust his working position multiple times, others may require him to sit and find a stable posture. Thus, the choice of the foot device or wearable has to be positioned with regard to the task at hand, considering aspects such as freedom of movement or how they would enable appropriate and comfortable feet-gestures. Even though TOOTHFAIRY v2's design was meant for the most common scenario in dentistry settings and followed a simplistic design which is familiar to dentistry professionals, it did not account for the size or materials of the foot pedal, which created the need for some users to adjust the pedal during task performance or to look at the pedal itself, to make sure they were pressing it correctly. Considering all of the above, we recommend: (i) In procedures that are performed in a sitting position and which consider using foot input, foot pedals should be regarded the input device of choice; (ii) In procedures that are performed standing and may require postural or positional adjustments, foot wearables are preferred. These should be comfortable and should not limit the natural movement of the body in anyway. While optical markers may require an intrusive setup and be subject to occlusions, inertial sensors can be considered as appropriate alternatives to detect simple tap gestures. Wireless Bluetooth communication between sensor and computer require a simpler yet reliable setup. To the best of our knowledge, such option has yet to be explored and requires further validation; (iii) Foot devices must be sterilizable and liquid resistant to surpass any difficulties concerning blood or lavage fluids. Since most foot-supported touchless medical image manipulation prototypes were not tested on real clinical contexts, so far this aspect has been overlooked; (iv) Foot devices should not require the user to look at them during the interaction process, considering appropriate sizes, materials and numbers of buttons.

6.3. Implications for the learning process

For medical dentists, innovation is often directly associated with a learning curve. Thus, the amount of effort involved in the learning process would need to be minimal for a novel technology to be adopted. Since foot pedals are already used to aid a number of tasks in clinical scenarios, extending foot input support seemed to be a natural way of creating new touchless interaction strategies that minimize the need to memorize several hand-gestures. This has been confirmed by our participants who preferred using TOOTHFAIRY over the hands-only condition. In spite of following a metaphor that is familiar to the domain (combining hands and feet), interaction design did not consider the existence of several types of foot pedal devices which are used even inside the same dental practice. As a result, our participants had to learn how to use a pedal in a different way than what they were used to on a daily basis, which required more effort than what was initially expected. Thus producing a negative impact that was not foreseen, in the form of unintended activations, namely unwanted window resizing, which were associated with the long click gesture. We then recommend: (i) Interaction design should privilege simple hand and feet gestures, similarly to what is discussed by [9, 5, 8]; (ii) The interface should provide a support or tutorial system that shows the user how to proceed and helps him to learn how to use the interface effectively. None of the works regarding foot-supported touchless interaction considered this possibility.

6.4. Implications for camera occlusions

Other concerns regard the ability of dealing with camera occlusions that would typically occur in a surgical setting. Kinect has been found to work best when faced frontally, registering a gradual decrease in performance with increasing view angles. Given Kinect's markerless skeleton tracking heavily depends on depth information, it is frequently affected by self-occlusion by other body parts and other objects in the scene [37]. As image manipulation may play a central role in guiding the clinical procedure, it is essential to guarantee the robustness of the gesture detection system. We recommend:

(i) Minimize occlusions while using depth cameras; system designers could either develop a network of cameras, merging the information captured, or place a single camera on the ceiling or at a high position. However, both of these options have limitations. Firstly, using several cameras would require a more complex and intrusive setup, which would not be usable in reduced workspaces. Also, placing the camera on higher levels would require accounting for a different acquisition angle, which may be problematic for gesture detection. Making sure users are at appropriate distances is also a concern. (ii) Another option to avoid occlusion is using wearables. Although the use of comfortable and non-limiting wearables may be feasible for feet input, using it for hand-gesture detection can be more challenging, especially because dental clinicians need to be able to hold surgical instruments. While [32] proposed combining hand wearables with floor sensors for touchless interaction with medical images, their evaluation was not performed with medical dentists, which limits the conclusions on its potential. Furthermore, in a scenario where dental chairs pose a serious challenge to detecting feet gestures, foot pedals would represent a more robust solution.

Finally, we should also consider aspects that are inherent to our target users. In Medical Dentistry, clinicians are well aware that the aesthetics of their work plays a major role on a patient's life [38, 39], which requires medical dentists to strive for excellence. Consequently, participants mentioned that it is only natural for them to take all the time they get to try to achieve the best results in every task they performed, affirming that *"Even in videogames, I only move on to the next level when I have the perfect (maximum) score"*. Ultimately, this suggested the time and effort taken to position each image would tend to be stretched to their maximum, which may explain the lack of significant differences in task completion times between both approaches.

6.5. Limitations of the user study

Our participant sample was limited to medical dentists with little experience in using 3D data. Our participants mostly dealt with 2D imagery, such as orthopantomography or digital radiography. Although we report results on a specific user population, they represent an important group when designing easy to use techniques. While we acknowledge that performance can be significantly different for expert users, the derive implications may still apply. Further research should replicate the user studies reported in the paper with proficient participants in 3D spatial manipulations.

It is also worth considering Kinect's accuracy and reliability. A study by Obdržálek et al. [37] reported typical errors of 5cm regarding pose estimation accuracy, with a variability of approximately 10cm for general postures, which may limit highly precise object manipulation and therefore enhance distances to target.

Finally, we address our research questions: (i) *What are the benefits and drawbacks of foot-supported 3D touchless image manipulation?* Foot-support has the potential to reduce the number of unintended activations and to achieve more accurate transformations, at the cost of a higher task completion time for new users; (ii) *How can these technologies be included in the daily clinical practices of medical dentists?* Interaction must be easy to learn and quick to respond, providing a seamless user experience. This includes considering confined cluttered spaces during interaction design; (iii) *Can foot device positioning positively affect medical image access?* Yes, in the sense that it affects user's mobility and the complexity of the gesture vocabulary; (iv) *Besides 2D data, is consulting with 3D data also relevant for the medical dentist?* Indeed, since 3D data enables an enhanced visualization of complex anatomical structures.

7. Conclusions and Future Work

Our work focused on touchless interaction techniques for medical image manipulation, based on 3D hand cursors supported by single-foot input, in the dental domain. We conducted a qualitative user study with medical dentists using a foot-supported gesture-based prototype as a design probe, which allowed us to investigate sitting and standing scenarios. This was followed by a quantitative study, to assess the impact of foot-supported interaction. To this end, participants performed tasks with a hands-only condition, used as a baseline, and the TOOTHFAIRY condition. Results showed statistically significant differences between both conditions regarding time, in two of the 2D tasks, and size (scale), in one of the 2D tasks, and most importantly, positioning precision, in one of the 3D tasks, and orientation, in two of the 3D tasks. Our findings indicate that foot-support can be a viable and better approach to both activation, and mode switching. Our approach was received positively by dental clinicians who are already familiar with using their feet while performing clinical procedures. In addition, 3D cursors were well received by medical dentists. We finish with guidelines for designing new foot-supported touchless medical image interaction that rely on 3D hand cursors. In the future, it would be interesting to consider how TOOTHFAIRY could be adapted to other medical specialties, so that our analysis can extend to other clinical scopes. Thus, future work will involve more diverse clinical teams participating in our study. Also, it would be relevant to investigate how training improves user performance throughout time, and to evaluate users' engagement and frustration during the learning process.

Acknowledgements

This work was partially supported by Fundação para a Ciência e a Tecnologia through grants UID/CEC/50021/2019, STREACKER UTAP-EXPL/CA/0065/2017, IT-MEDEX PTDC/EEI-SII/6038/2014, SFRH/BPD/97449/2013 and SFRH/BD/136212/2018. Authors would like to acknowledge the invaluable help of Dr. Pedro Rodrigues and his students at Instituto Universitário Egas Moniz for their willingness to collaborate throughout this work.

References

References

[1] G. M. Rosa, M. L. Elizondo, Use of a gesture user interface as a touchless image navigation system in dental surgery: Case series report, Imaging science in dentistry 44 (2) (2014) 155–160. doi:10.5624/isd.2014.44.2.155.

[2] R. Marklin, K. Cherney, Working postures of dentists and dental hygienists, Journal of the California Dental Association 33 (2) (2005) 133–136.

[3] W. Wan Hassan, N. Abu Kassim, A. Jhawar, N. M. Shurkri, N. Kamarul Baharin, C. Chan, User acceptance of a touchless sterile system to control virtual orthodontic study models, American Journal of Orthodontics and Dentofacial Orthopedics 149 (4) (2016) 567–578. doi:10.1016/j.ajodo.2015.10.018.

[4] M. A. Mora, D. L. Chenin, R. M. Arce, Software tools and surgical guides in dental-implant-guided surgery, Dental Clinics of North America 58 (3) (2014) 597 – 626, cone Beam Computed Tomography: From Capture to Reporting. doi:https://doi.org/10.1016/j.cden.2014.04.001.

[5] K. O'Hara, G. Gonzalez, A. Sellen, G. Penney, A. Varnavas, H. Mentis, A. Criminisi, R. Corish, M. Rouncefield, N. Dastur, T. Carrell, Touchless interaction in

surgery, Commun. ACM 57 (1) (2014) 70–77. doi:10.1145/2541883.2541899.

[6] D. Lopes, P. Parreira, S. Paulo, V. Nunes, P. Rego, M. Neves, P. Rodrigues, J. Jorge, On the utility of 3d hand cursors to explore medical volume datasets with a touchless interface, Journal of Biomedical Informatics 72 (Supplement C) (2017) 140 – 149. doi:https://doi.org/10.1016/j.jbi.2017.07.009.

[7] A. Mewes, B. Hensen, F. Wacker, C. Hansen, Touchless interaction with software in interventional radiology and surgery: a systematic literature review, International Journal of Computer Assisted Radiology and Surgery 12 (2) (2017) 291–305. doi:10.1007/s11548-016-1480-6.

[8] S. Cronin, G. Doherty, Touchless computer interfaces in hospitals: A review, Health Informatics Journaldoi:10.1177/1460458217748342.

[9] J. Schöning, F. Daiber, A. Krüger, M. Rohs, Using hands and feet to navigate and manipulate spatial data, in: CHI '09 Extended Abstracts on Human Factors in Computing Systems, CHI EA '09, ACM, New York, NY, USA, 2009, pp. 4663–4668. doi:10.1145/1520340.1520717.

[10] C. Holz, A. Wilson, Data miming: Inferring spatial object descriptions from human gesture, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM, New York, NY, USA, 2011, pp. 811–820. doi:10.1145/1978942.1979060.

[11] Vinayak, S. Murugappan, H. Liu, K. Ramani, Shape-it-up: Hand gesture based creative expression of 3d shapes using intelligent generalized cylinders, Computer-Aided Design 45 (2) (2013) 277 – 287, solid and Physical Modeling 2012. doi:https://doi.org/10.1016/j.cad.2012.10.011.

[12] F. Göbel, K. Klamka, A. Siegel, S. Vogt, S. Stellmach, R. Dachselt, Gaze-supported foot interaction in zoomable information spaces, in: CHI '13 Extended Abstracts on Human Factors in Computing Systems, CHI EA '13, ACM, New York, NY, USA, 2013, pp. 3059–3062. doi:10.1145/2468356.2479610.

[13] R. Johnson, K. O'Hara, A. Sellen, C. Cousins, A. Criminisi, Exploring the potential for touchless interaction in image-guided interventional radiology, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11, ACM, New York, NY, USA, 2011, pp. 3323–3332. doi:10.1145/1978942.1979436.

[14] A. V. Reinschluessel, J. Teuber, M. Herrlich, J. Bissel, M. van Eikeren, J. Ganser, F. Koeller, F. Kollasch, T. Mildner, L. Raimondo, L. Reisig, M. Ruedel, D. Thieme, T. Vahl, G. Zachmann, R. Malaka, Virtual reality for user-centered design and evaluation of touch-free interaction techniques for navigating medical images in the operating room, in: Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '17, ACM, New York, NY, USA, 2017, pp. 2001–2009. doi:10.1145/3027063.3053173.

[15] M. Jacob, J. Wachs, Context-based hand gesture recognition for the operating room, Pattern Recognition Letters 36 (2014) 196 – 203. doi:https://doi.org/10.1016/j.patrec.2013.05.024.

[16] D. B. Douglas, R. E. Douglas, C. Wilke, D. Gibson, J. Boone, M. Wintermark, A systematic review of 3d cursor in the medical literature, AIMS Electronics and Electrical Engineering 2 (ElectronEng-02-01-001) (2018) 1. doi:10.3934/ElectrEng.2018.1.1.

[17] M. Ma, P. Fallavollita, S. Habert, S. Weidert, N. Navab, Device-and system-independent personal touchless user interface for operating rooms, International Journal of Computer Assisted Radiology and Surgery 11 (6) (2016) 853–861. doi:10.1007/s11548-016-1375-6.

[18] A. Meneses, B. Rocha, A. Nogueira, P. Mattos, R. Campos, G. Guimaraes, S. Zequi, 944 a new gesture-controlled tool using three-dimensional reconstruction of renovascular-collecting system-tumor anatomy to assist navigation of kidney during "zero ischemia" minimally invasive nephron sparing surgery in high complex renal cancer, European Urology Supplements 15 (3) (2016) e944.

[19] J. Hettig, P. Saalfeld, M. Luz, M. Becker, M. Skalej, C. Hansen, Comparison of gesture and conventional interaction techniques for interventional neuroradiology, International journal of computer assisted radiology and surgery 12 (9) (2017) 1643–1653. doi:10.1007/s11548-017-1523-7.

[20] B. Hatscher, M. Luz, L. Nacke, N. Elkmann, V. Müller, C. Hansen, Gazetap: Towards hands-free interaction in the operating room, in: Proceedings of the 19th ACM International Conference on Multimodal Interaction, ICMI 2017, ACM, New York, NY, USA, 2017, pp. 243–251. doi:10.1145/3136755.3136759.

[21] D. Black, M. Unger, N. Fischer, R. Kikinis, H. Hahn, T. Neumuth, B. Glaser, Auditory display as feedback for a novel eye-tracking system for sterile operating room interaction, International journal of computer assisted radiology and surgery 13 (1) (2018) 37–45. doi:10.1007/s11548-017-1677-3.

[22] T. Pakkanen, R. Raisamo, Appropriateness of foot interaction for non-accurate spatial tasks, in: CHI '04 Extended Abstracts on Human Factors in Computing Systems, CHI EA '04, ACM, New York, NY, USA, 2004, pp. 1123–1126. doi:10.1145/985921.986004.

[23] A. Rovers, H. van Essen, Guidelines for haptic interpersonal communication applications: an exploration of foot interaction styles, Virtual Reality 9 (2) (2006) 177–191. doi:10.1007/s10055-005-0016-0.

[24] D. Dearman, A. Karlson, B. Meyers, B. Bederson, Multi-modal text entry and selection on a mobile device, in: Proceedings of Graphics Interface 2010, GI '10, Canadian Information Processing Society, Toronto, Ont., Canada, Canada, 2010, pp. 19–26.

[25] J. Scott, D. Dearman, K. Yatani, K. Truong, Sensing foot gestures from the pocket, in: Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology, UIST '10, ACM, New York, NY, USA, 2010, pp. 199–208. doi:10.1145/1866029.1866063.

[26] T. Augsten, K. Kaefer, R. Meusel, C. Fetzer, D. Kanitz, T. Stoff, T. Becker, C. Holz, P. Baudisch, Multitoe: High-precision interaction with back-projected floors based on high-resolution multi-touch input, in: Proceedings of the 23Nd Annual ACM Symposium on User Interface Software and Technology, UIST '10, ACM, New York, NY, USA, 2010, pp. 209–218. doi:10.1145/1866029.1866064.

[27] A. Crossan, S. Brewster, A. Ng, Foot tapping for mobile interaction, in: Proceedings of the 24th BCS Interaction Specialist Group Conference, BCS '10, British Computer Society, Swinton, UK, UK, 2010, pp. 418–422.

[28] G. Rivière, N. Couture, P. Reuter, The activation of modality in virtual objects assembly, Journal on Multimodal User Interfaces 3 (3) (2010) 189–196. doi:10.1007/s12193-010-0038-0.

[29] R. Balakrishnan, G. Fitzmaurice, G. Kurtenbach, K. Singh, Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip, in: Proceedings of the 1999 Symposium on Interactive 3D Graphics, I3D '99, ACM, New York, NY, USA, 1999, pp. 111–118. doi:10.1145/300523.300536.

[30] K. Zhong, F. Tian, H. Wang, Foot menu: Using heel rotation information for menu selection, in: 2011 15th Annual International Symposium on Wearable Computers, 2011, pp. 115–116. doi:10.1109/ISWC.2011.10.

[31] A. L. Simeone, E. Velloso, J. Alexander, H. Gellersen, Feet movement in desktop 3d interaction, in: 2014 IEEE Symposium on 3D User Interfaces (3DUI), 2014, pp. 71–74. doi:10.1109/3DUI.2014.6798845.

[32] S. Jalaliniya, J. Smith, M. Sousa, L. Büthe, T. Pederson, Touch-less interaction with medical images using hand & foot gestures, in: Proceedings of the 2013 ACM Conference on Pervasive and Ubiquitous Computing Adjunct Publication, UbiComp '13 Adjunct, ACM, New York, NY, USA, 2013, pp. 1265–1274. doi:10.1145/2494091.2497332.

[33] G. Pearson, M. Weiser, Of moles and men: The design of foot controls for workstations, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '86, ACM, New York, NY, USA, 1986, pp. 333–339. doi:10.1145/22627.22392.

[34] W. Saunders, D. Vogel, The performance of indirect foot pointing using discrete taps and kicks while standing, in: Proceedings of Graphics Interface 2015, GI 2015, Canadian Human-Computer Communications Society, Toronto, Ontario, Canada, 2015, pp. 265–272. doi:10.20380/GI2015.34.

[35] S. F. Paulo, M. Belo, R. K. dos Anjos, J. A. Jorge, D. S. Lopes, Volume and Surface Rendering of 3D Medical Datasets in Unity®, MICCAI Educational Challenge 2017-2018.

[36] D. Mendes, F. Relvas, A. Ferreira, J. Jorge, The benefits of dof separation in mid-air 3d object manipulation, in: Proceedings of the 22Nd ACM Conference on Virtual Reality Software and Technology, VRST '16, ACM, New York, NY, USA, 2016, pp. 261–268. doi:10.1145/2993369.2993396.

[37] Š. Obdržálek, G. Kurillo, F. Ofli, R. Bajcsy, E. Seto, H. Jimison, M. Pavel, Accuracy and robustness of kinect pose estimation in the context of coaching of elderly population, in: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, 2012, pp. 1188–1193.

[38] L. Davis, P. Ashworth, L. Spriggs, Psychological effects of aesthetic dental treatment, Journal of dentistry 26 (7) (1998) 547–554.

[39] D. M. Roessler, Complete denture success for patients and dentists, International dental journal 53 (S5) (2003) 340–345.

Appendix

Table 1: Questionnaire results comparing standing and sitting, using an optical marker (P1) and a foot pedal (P2), respectively: UT (Unconstrained Translation), CT (Constrained Translation), UR (Unconstrained Rotation), CR (Constrained Rotation), S (Scale), SN (Slice Navigation), VS (Viewport Selection), VR (Viewport Resizing). Median (Interquartile Range). *= statistically significant ($p \leq 0.05$)

	Usefulness		Execution		Memorability		Usability	
	P1	P2	P1	P2	P1	P2	P1	P2
UT	6.00 (1.00)	6.00 (1.00)	6.00 (2.00)	5.00 (1.00)	**5.00** (0.75)	**5.50*** (2.00)	5.00 (2.00)	5.00 (1.00)
CT	6.00 (1.00)	6.00 (1.25)	5.50 (2.00)	5.00 (1.00)	**4.50** (1.25)	**5.00*** (2.00)	5.00 (2.00)	5.00 (1.00)
UR	6.00 (1.25)	6.00 (1.00)	5.50 (1.25)	5.00 (1.00)	**5.00** (1.00)	**5.50*** (2.00)	5.00 (2.00)	5.00 (1.00)
CR	6.00 (1.25)	6.00 (1.00)	5.00 (1.25)	5.00 (1.00)	**4.50** (1.00)	**5.00*** (2.00)	5.00 (2.00)	5.00 (1.00)
S	6.00 (1.00)	6.00 (1.00)	6.00 (1.00)	5.00 (1.00)	5.00 (2.00)	6.00 (2.00)	5.00 (2.00)	5.00 (1.00)
SN	5.50 (1.25)	6.00 (1.25)	5.50 (1.25)	5.00 (1.25)	5.00 (2.00)	6.00 (1.00)	5.00 (1.25)	5.00 (1.00)
VS	6.00 (1.25)	6.00 (2.00)	5.50 (2.00)	5.50 (1.25)	5.00 (2.00)	6.00 (1.25)	5.00 (2.00)	5.50 (1.00)
VR	5.50 (1.25)	6.00 (2.00)	5.50 (2.00)	5.50 (1.25)	5.00 (2.00)	6.00 (1.00)	5.00 (1.00)	5.00 (1.25)

Table 2: Task A results comparing baseline and TOOTHFAIRY conditions. Mean (standard deviation).*= statistically significant ($p \leq 0.05$)

Measurements	Baseline	TOOTHFAIRY	Z	p
slice distance (n slices)	.74 (.93)	.45 (.69)	-1.387	.165
distance (cm)	.09 (.39)	.00 (.00)	-1.000	.317
distance X (cm)	.08(.35)	.00 (.00)	-1.342	.180
distance Y (cm)	.10 (.31)	.00 (0.00)	-1.342	.180
scale error (factor)	**.24 (.44)**	**.00* (.00)**	-2.366	**.018**
time (s)	49.75 (29.50)	52.76 (29.10)	-.504	.614

Table 3: Task B results comparing baseline and TOOTHFAIRY conditions. Mean (standard deviation).*= statistically significant ($p \leq 0.05$)

Measurements	Baseline	TOOTHFAIRY	Z	p
slice distance (n slices)	.00 (.00)	.00 (.00)	.000	1.000
distance (cm)	2.51 (1.49)	2.78 (1.48)	-.276	.783
distance X (cm)	.83 (.81)	1.08 (.84)	-.784	.433
distance Y (cm)	1.84 (1.41)	2.21 (1.45)	-.633	.527
scale error (factor)	.39 (.32)	.33 (.29)	-.226	.821
time (s)	**30.20 (17.23)**	**41.83* (20.57)**	-2.576	**.010**

Table 4: Task C results comparing baseline and TOOTHFAIRY conditions. Mean (standard deviation).*= statistically significant ($p \leq 0.05$)

Measurements	Baseline	TOOTHFAIRY	Z	p
slice distance (n slices)	1.31 (1.45)	3.31 (4.60)	-.598	.550
distance (cm)	2.40 (1.01)	2.08 (1.41)	-.971	.332
distance X (cm)	1.16 (.99)	1.45 (.99)	-1.130	.259
distance Y (cm)	**1.83 (1.20)**	**1.06* (.73)**	-2.415	**.016**
scale error (factor)	.57 (.44)	.38 (.31)	-1.248	.212
time (s)	**60.31 (28.98)**	**93.42* (30.66)**	-2.678	**.007**

Table 5: Task D results comparing baseline and TOOTHFAIRY conditions. Mean (standard deviation).*= statistically significant ($p \leq 0.05$)

Measurements	Baseline	TOOTHFAIRY	Z	p
distance (cm)	4.41 (4.09)	3.20 (3.34)	-1.546	.122
distance X (cm)	2.32 (2.10)	1.40 (1.46)	-1.475	.140
distance Y (cm)	2.37 (3.90)	1.36 (2.04)	-.664	.507
rotation error (°)	10.12 (9.14)	8.02 (9.23)	-1.138	.255
rotation error X (°)	2.14 (3.46)	2.55 (3.30)	-.259	.796
rotation error Y (°)	7.05 (8.86)	5.71 (7.79)	-.776	.438
rotation error Z (°)	**3.97 (4.96)**	**1.25* (2.09)**	-1.988	**.047**
scale error (factor)	.17 (.29)	.20 (.30)	-.220	.826
time (s)	79.84 (37.60)	87.13 (31.79)	-.365	.715

Table 6: Task E results comparing baseline and TOOTHFAIRY conditions. Mean (standard deviation).*= statistically significant ($p \leq 0.05$)

Measurements	Baseline	TOOTHFAIRY	Z	p
distance (cm)	6.92 (5.87)	5.96 (4.09)	-.207	.836
distance X (cm)	5.27 (5.42)	4.48 (4.39)	-.747	.455
distance Y (cm)	2.90 (2.20)	2.45 (2.27)	-.596	.551
rotation error (°)	12.65 (13.06)	10.14 (9.52)	-.362	.717
rotation error X (°)	10.00 (12.72)	9.46 (10.26)	-.226	.821
rotation error Y (°)	1.87 (2.50)	3.04 (3.67)	-.031	.975
rotation error Z (°)	.77 (1.09)	1.28 (1.54)	-.260	.795
scale error (factor)	.02 (.032)	.03 (.04)	-.847	.397
time (s)	78.90 (33.33)	93.79 (27.60)	-1.380	.168

Table 7: Task F results comparing baseline and TOOTHFAIRY conditions. Mean (standard deviation).*= statistically significant ($p \leq 0.05$)

Measurements	Baseline	TOOTHFAIRY	Z	p
distance (cm)	10.20 (7.05)	9.03 (7.45)	-.821	.411
distance X (cm)	7.56 (5.34)	4.61 (2.61)	-1.586	.113
distance Y (cm)	3.09 (3.17)	5.45 (6.81)	-.781	.435
rotation error (°)	13.30 (9.59)	10.54 (10.38)	-1.681	.093
rotation error X (°)	5.91 (4.01)	7.79 (9.55)	-.517	.605
rotation error Y (°)	8.00 (8.36)	2.19 (2.62)	-1.655	.098
rotation error Z (°)	**1.89 (2.41)**	**.66* (.99)**	-2.069	**.039**
scale error (factor)	.21 (.19)	.21 (.21)	-.226	.821
time (s)	87.58 (39.04)	94.71 (31.22)	-.617	.537