ABSTRACT
This paper studies the capabilities, limitations and potential of combining a vision-based system with EMG sensors for freehand interaction in mixed reality environments. We present the design and implementation of our system using the Hololens and Myo armband, conduct a preliminary user study with 15 participants to study the usability of our model and discuss the advantages, potentials and limitations of this approach, and discuss our findings and its implications for the design of user interfaces with a similar hardware setup.

We show that the flexibility of interaction in our proposal has positive effects on the user performance for the completion of a complex user task, although measured user performance for the individual gestures was worse on average than the performance obtained for the gestures supported by the standard Holotoolkit. One can conclude that the presented interaction paradigm has great potential for future use in mixed reality, but it still has some limitations regarding robustness and ergonomics that must be addressed for a better user acceptance and broader public adoption.

Author Keywords
Freehand interaction; mid-air gestures; EMG interfaces; multimodal interaction

INTRODUCTION
As the adoption of mixed reality increases rapidly, the need to develop and improve more natural and intuitive interaction models gets more and more relevant. Consequently, new and more robust paradigms for improved freehand gestures as a natural form of interaction require further development and evaluation. A traditional approach to enable such interaction is the usage of vision-based systems. However, other alternatives, such as wearables based on EMG data are often overlooked even though they offer promising opportunities for overcoming some of the limitations inherent to the vision-based systems and help to provide a more robust and flexible interaction that can be ultimately more intuitive.

With the development of more advanced Mixed Reality platforms, such as the Microsoft Hololens [11], the users do not only expect the ability to see virtual elements in coherence with the real space, but they also want to interact directly with such elements in a more natural way. They expect to approach a virtual model on top of a real table, touch it and manipulate it as they would do with any other real object in their environment.

Most platforms rely only on their vision systems to provide freehand interaction. In order to guarantee an acceptable robustness and performance in real time, the result is often a very limited set of gestures that may not be the best fit for the different tasks that an user expects to accomplish in a mixed reality environment. In the case of the Hololens, the possible gestures for interaction are essentially two (bloom and air-tap) plus the tracking of hands as included in the latest software update for the first generation of the Hololens [11].

On the other hand, some wearable devices such as smartwatches, that provide orientation data from inertial measurement units (IMU) have been used as interfaces in different applications and gained significant attention in HCI [39]. However, wearables based on electromyography (EMG) sensors or surface electromyography (sEMG) are still not explored enough. They have been long studied before but mostly for medical applications only, yet their potential in the area of Human-Computer Interaction and mixed reality still requires further exploration and evaluation. The Myo armband by Thalmic Labs\(^1\) has been so far the first commercial wearable and the most complete solution in this category [44] containing an array of eight surface EMG sensors and a 9-DOF Inertial Measurement Unit, plus a proprietary firmware that offers the recognition five different gestures. Therefore, in this work we propose a new interaction model that takes advantage of the Myo armband technology in order to overcome limitations inherent to the vision system of the Hololens.

Since the Myo armband does not require the gestures to be made in front of a camera, it poses as a suitable complement to the Hololens: it offers an alternative for interaction in conditions where the Hololens would not be able to respond, such as occlusion of the hands, or in the case where the hands are out of its field of view (FOV) [20, 37]. In this work we also show how expanding the set of gestures by integrating the Myo gestures in combination with the ones from the Hololens provides higher flexibility and more intuitive interaction possibilities.

We have applied our interaction model for the case of 3D object manipulation given that many tasks in virtual and augmented reality involves the ability to manipulate 3D models in the environment [21]. For this purpose, we support the basic transformations for 3D rigid bodies: translation, rotation and scaling.

We describe our use-case scenario used for the evaluation of our system and we show that: (i) our interaction model is more intuitive than the standard model based on the vision system of the Hololens, (ii) that providing a higher flexibility reflects in a better performance overall given a longer task (iii) and that although the performance of new introduced gestures based on EMG were less efficient than those based on the vision system, the preference of the users still favored the intuitiveness when allowed to chose. We conclude with

---

\(^1\)The Myo armband was discontinued in 2018 however the existing units are still being supported. https://bit.ly/2OErQRN
a discussion of advantages, potentials and limitations of this interaction paradigm as well as of the EMG as a growing technology in interaction systems.

RELATED WORK
Freehand interaction in Mixed reality
Freehand interaction and mid-air gestures are often recommended when designing natural interfaces [2, 7] since they are intuitive, as they can relate better with interactions with the physical world, and are less cumbersome as they don’t need controllers to be hold [15]. When used in Mixed-Reality systems, this interaction style is most commonly supported by vision systems, either with visible-light cameras or infrared cameras [7]. Among the methods for visual classification of hand gestures the usage of deep-learning models are reporting the most outstanding results [19, 26, 41], and among those, the methods incorporating fusion techniques combined with mesh-smoothing optimizations report the best results [42], indicating a strong advantage for hardware setups with multiple cameras of the same or different type. However, an ever present constraint of the state-of-the-methods is the need of considerably high memory and processing capabilities, which is not the case for the main platforms for Mixed-Reality running on limited embedded cores.

The first generation of the Microsoft Hololens for instance handles these constraints by providing a very limited set of gestures for interaction that are easily classified and therefore provide higher reliability and robustness for the user experience [9].

Previous work has tried to enhance the inherent interaction model of the Hololens by combining with other user interfaces such as smartwatches, or most commonly the Leap motion. However, none of the recent work have studied the fusion with EMG wearable solutions.

EMG-based interaction
We have found multiple examples of EMG-based interfaces [16, 31, 33, 48] including Myo Armband that show an increasing interest in such technology for HCI, although still in a premature phase. The Myo armband by Thalmic Labs is the most popular wearable of its kind for commercial use given its ergonomic and adaptive design, and the pre-trained model that already offers the possibility to recognize five different gestures as shown in figure 1. Although the Myo has been used as interface for different systems [13, 27, 35, 39, 46] its usage in the field of Interaction for Mixed-reality is still not studied enough. For instance, we found very few examples where the hardware setup combines the Hololens (being the most complete headset in this category) and the Myo on the same platform, but none of them is oriented to improve the interaction paradigm for Mixed Reality environments. In the case of wekit [16, 38], the Myo has been used as an input device to collect data as part of their expert elicitation system, rather than as an user interface. In another example [46] the Myo is used to control the velocity of a mobile robot; however, the interaction is rather limited and independent from the Mixed Reality system itself, as it only acts as a binary command to increase or reduce the velocity of the robot. In a most recent work we found that the Myo Armband together with the Hololens has been used as a training system for amputees prostheses [39, 44], however, the mapping of gestures during such training by its nature is not designed to be intuitive, in this case the mapped input do not correspond to the visual feedback in the augmented reality environment. For example, the muscle activity corresponding to a fist in the forearm is mapped in the environment as a fingerpoint action. [8]

Due to the nature of the sEMG technology it is important to notice that the muscles that can be measured are mostly in charge of the flexion, extension, adduction and abduction movements of the hands, plus the pronation and supination of the forearm, rather than the control of individual fingers, which are mainly controlled by muscles in the intermediate and deep layer [44]. Only the Flexor digitorum superficialis is directly connected to the phalanges in the hand, but even this muscles is barely present on the superficial layer and is often considered only in the middle layer of the forearm muscles [14, 28, 29]. The intermediate and deep layer of muscles are where the main muscles that controls the movements of the phalanges are present, unfortunately for the surface EMG sensors, the signals from such muscles are hardly detected in the superficial layer. This poses additional challenges for the recognition of gestures and movements based solely on EMG data, although recent research has shown interesting results using deep learning techniques [1, 3, 8, 10, 12, 18, 25, 32, 34, 47], however this is still work in progress and by the time being we have to consider that the number and type of gestures recognized in a reliable manner are rather limited for practical applications.

In the future however, we should also consider that implantable solutions IEMS such as [22, 23, 40, 45] are expected to be easier accessed and will probably be more accepted by users and society as the technology advances and gets more attention by researchers in HCI.

DESIGN AND IMPLEMENTATION
Our interaction model is designed to fulfill three main objectives. First, to provide a fluid interaction even when the model manipulated is outside of the field of view of the Hololens. Second, to reduce users fatigue by identifying and integrating the most comfortable arm and hand positions for extended use. Third, to provide a bigger set of gestures in the context of 3D Objects manipulation that ultimately allow the user the choice for the most intuitive interaction.

Design considerations
Our proposed interaction takes into account design guidelines from previous work in hand-gestures design [2, 6, 15, 17, 30, 36] as well as previous studies where the fatigue caused by some hand and forearm postures has been evaluated [4, 17, 36]. In particular, with the usage of the armband as additional interaction we are minimizing effects like the monkey syndrome identified in previous studies.

Also, in order to provide a natural interaction we also take into account some daily natural actions, such as "grabbing and object" (fist) or "turning a screwdriver" (forearm pronation) that can be mapped into our interaction model.
We have decided to include only a sub-set of Myo gestures to complement the Hololens interaction. These are the fist, the spread fingers and the double-tap gesture as shown in figure 1. The wave-in and wave-out gestures were discarded since they have been identified in a previous study as some of the most cumbersome gestures that can cause fatigue for extended usage [36], while the fist is the gesture chosen to be held on for most of the interaction with the myo armband for two main reasons, the first one being its identification as one of the most comfortable gestures in the same study, and second because for the case of 3D Objects manipulation, it emulates the "grabbing" gesture that any user is already familiar with in a real environment, making it the most natural choice for such interaction. Finally, the spread-fingers and the double-tap are used only for brief periods of time. In addition, we use the orientation vector provided by the inertial unit in the armband in combination with each of the gestures while transforming the 3D Objects in the scene.

Interaction Model
In our model, we reuse the interaction mapping from the Two Hand Manipulatable script of the Mixed-Reality Toolkit also called Holotoolkit [24] and integrate in parallel our Myo-supported interaction model as shown in figure 2, thus providing a permanent alternative of interaction for each transformation that responds better to the personal preferences of the users.

In the standard Holotoolkit interaction the gestures use both hands for the case of rotation and scaling, and one or two hand for translating. In all cases the air-tap gesture is used.

In the proposed model supported by the Myo armband we have favored the fist gesture for its usage in all the transformations as it has been identified in the literature as one of the less fatiguing positions [36]. The final interaction model is summarized in table 1.

<table>
<thead>
<tr>
<th>Operation</th>
<th>Hololens-only</th>
<th>Myo-supported</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate</td>
<td>hold air tap + release</td>
<td>hold fist and release</td>
</tr>
<tr>
<td>Rotate</td>
<td>two-hands air tap</td>
<td>fist and rotate forearm</td>
</tr>
<tr>
<td>Scale</td>
<td>two-hands air tap</td>
<td>fist and spread fingers</td>
</tr>
</tbody>
</table>

PERLIMINAR USER STUDY
Use Case Scenario
The implemented system supports the basic transformations for rigid objects in a 3D scenario, i.e., translation, rotation and scaling. As one example for evaluation, we select a scenario where we display a simple model of a rocket, consisting of four pieces: engine mount, body tube, body cap and nose cone as illustrated in figure 4.

The experiment was conducted in a room with dimensions of 3m x 4m where the 3D objects to be manipulated were distributed in a semi random way. The room included two tables separated by 1.5m, where some targets were marked with regular tape as indication for the experiment (as shown in figure 5).

Experiment
The goal of this experiment is to determine the advantages and limitations of our proposal compared to the standard interaction model of the Holotoolkit, i.e. Hololens only. To do so we evaluate some aspects of usability such as user performance and acceptance for both the individual gestures and for the completion of a more complex task defined in advance in the context of 3D Object manipulations.
The experiment procedure included three stages: (i) introduction and learning of the interaction models, (ii) individual gesture evaluation and (iii) free interaction evaluation. For the second and third stage we measured the user performance as the completion time for each task.

And finally, at the end of the experiment the participants were asked to fill out a questionnaire to establish the demographics of the participants and assess the general user acceptance of the interaction model. The questionnaire includes questions, such as gender, age, background and previous experience with XR devices and EMG devices. We then included some questions that specifically evaluated the level of comfort perceived for each of the gestures included in our interaction model. Participants were asked to answer the questions on a seven-point Likert scale (Strongly Disagree, Disagree, Somewhat Disagree, Neutral, Somewhat Agree, Agree, Strongly Agree). We further included some open questions concerning the preferred gestures for each operation and the reasons for it, and also the participants were invited to provide suggestions regarding the interaction model and other possible desirable additional gestures.

Introduction and learning of the interaction models
During this familiarization stage each participant was introduced to the technology and interaction model, including a pre-calibration procedure for the Myo armband. We asked participants to perform each of the possible gestures applied to a single object in the 3D scene (a regular cone representing the top of the rocket model). They were allowed to experiment freely with the system for about five minutes in order to familiarize themselves with the hardware, gestures and expected response of the system.

Gesture evaluation
In the second stage, we measured completion times for each of the gestures providing the following instructions:

- Rotate the engine mount (rocket base) 360 degrees along the z-axis as defined by a user’s line of sight.
- Translate the body tube (main cylinder) to a designated area.
- Up-scale the body cap (small cylinder) in size and down-scale it back to its original size.

For each task we asked participants to repeat the procedure, one time using only the standard Hololens-based interaction (air tap & double hands) and another time using only the gestures supported by the Myo-armband (double-tap, fist and extend-fingers). The order of such tasks was not specifically given so the users could decide any order to perform each of them.

Free interaction task
For the final stage, we asked participants to assemble the complete rocket model using the floating parts spread around the environment. We placed a copy of the rocket already assembled as reference on top of a table and asked participants to build their own model next to it on a marked spot. For this task the participants used any gesture of their choice. We recorded their choices.

Figure 5 shows an example of one of the finished tasks for the third stage of the experiment, where users were asked to assemble the rocket in a free manner, i.e., allowing them to choose which gesture to use at any given time.

USER STUDY RESULTS
The user study was conducted with 15 participants (12 males and 3 females), between the ages of 23 and 32 years. Their fields of study and expertise varied from software engineering and intelligent systems to physical biology, civil engineering and commercial vehicle technology. Five of the participants had previously used an AR or VR device (Oculus Rift). The others had no prior experience with extended reality systems. No participant had used the Hololens before. Only one participant had used the Myo armband before.

When comparing the completion time for each gesture (figure 6) we observed that in general the gestures supported as part of the Holotoolkit were more efficient, and that those supported by the MYO armband were performed slower in average but
also showed a higher variance between users suggesting that wearable interfaces based EMG technology is more susceptible to the particular characteristics of each user and that a more thorough calibration process is required in order to get more consistent results among users, since factors such as the fat tissue and muscle strength play a significant role in the final robustness observed.

However, for the third stage of evaluation we observed an improvement in the completion times for the overall task when using the combination of both interaction modalities according to the users’ preferences (figure 7).

In regards to the user acceptance of the system we analyze the responses in perceived comfort for each operation as well as the chosen gestures when performing the task of free interaction.

We observed that the Hololens standard interaction supported by the Holotoolkit was preferred for the rotation and translation actions, but the scaling operation done with the Myo was perceived in general more comfortable over scaling with the Hololens.

As observed in figure 9, most participants preferred to perform rotations with the Myo than with the Hololens due to the degree of intuitive use. Rotating with the Hololens required participants to use both hands, and sometimes the Hololens did not recognize both hands.

**Empirical observations and users’ feedback**

Regarding the observations of the users during the evaluation we collected several remarks. Some participants had problems with the weight of the Hololens, making its use uncomfortable for them. The Myo armband was “too tight” for some participants, leaving indentations on their arms. Although the Myo did not make them uncomfortable, participants expressed relief when it was taken off. The application, for the most part, worked as intended. Despite some unexpected behavior seen during interaction, users were able to recover and continue with a task. Participants were able to perform most of the gestures easily and found them to be natural and intuitive.

Scaling was executed more “smoothly,” requiring the use of one hand only, a characteristic that users appreciated. The Hololens required participants to use both hands in front of the camera to scale; participants found it difficult to execute tasks correctly.

Although the participants liked the idea of using the Myo to translate objects, the inaccuracy of translation made them prefer the use of the Hololens to perform translation. Some participants used the Myo armband to translate objects over larger distances and the Hololens to fine-tune final object placement. It is important to point out that during the evaluation some users experienced an unexpected artifact during interaction with the Myo, e.g., a temporary shaking movement of the 3D object being manipulated when releasing a fist gesture, as a result of noise present in the IMU data when switching to a gesture. We noticed that these artifact were more severe for some users. We believe that this effect is due to noise in the signals generated by a device. However, we consider that this artifact did not affect our results, as only time was measured.
and the criteria for a completed task were established when a user was satisfied with the final position, rotation phase and size of each of the 3D objects.

In the case of the Myo gestures, many participants expressed difficulty to perform the Double Tap gesture while for some others it worked as intended every time. Although it was not our focus to test the robustness for the recognition of each gesture, we observed that the disparity of performance between users when using the Myo armband was reflected in a higher deviation standard in the completion times for individual gestures. Participants provided suggestions regarding desirable additional gestures to be supported. For example, some users would like to use the Wave In and Wave Out gestures of the Myo to move objects sideways, in addition to using just the existing translation capability. Some users indicated that they would prefer a gesture different from Double Tap to switch modes for Myo manipulations as that gesture was not easily recognized.

In addition, some of the users included suggestions in regards to the visual design as they desired some additional the feedback during interaction. One participant suggested to incorporate visual feedback, such as a mild flash or changing background color. Additionally, as the Myo armband can vibrate, we could use that capability as well to inform a user that she has performed an action or is beginning to perform one. Some of the participants had no intention of using the Hololens with the Myo armband in the future. Since our participants had diverse backgrounds, some of them are likely to have no need to use Mixed Reality devices as part of their profession.

Another important remark that can be observed in figure 7 that most of the participants took less time to complete the task during the free interaction, i.e., when using both, the Myo armband and the Hololens-based gestures

**Discussion**

In our experiment, we investigated the the user acceptance towards the usage of the proposed interaction model in comparison with the standard interaction capabilities of a vision-based system. We evaluated the attitude of the users towards our multimodal setup and tested the difference of performance when using different sets of gestures in order to complete a task fairly common in mixed reality environments, e.g., applying rigid transformations to 3D Objects in the scene.

We can conclude that users generally felt positive about the interaction paradigm implemented in our system and its potential. Our prototype interaction system was sufficient to demonstrate intuitive and adequate interaction for 3D object manipulations. Among some interesting remarks we found that although some gestures supported by the MYO armband proved to be less efficient according to the measured performance 6, the users still chose to use them when facing the free interaction task. Also important is that even though the Hololens standard gestures performed overall better than the MYO supported gestures, when using both systems in combination the completion of the assembly task was more efficient, proving then that the flexibility of interaction represented a significant advantage in the multimodal interaction model.

**CONCLUSIONS AND FUTURE WORK**

We presented the design and implementation of an interaction model combining the advantages of a vision-based system, the most commonly one used for freehand interaction, with a wearable interface based on EMG sensors and inertial measurements. Our model can overcome several of the limitations typical for vision-based systems by allowing a fluid interaction even when hands are out of the camera’s field of view (FOV) or occluded by other objects. Our approach is aiming at reducing fatigue of the upper arm as gestures do not need to be sustained in a high position to lie inside the FOV.

Our model offers a high degree of flexibility and intuitiveness for interaction in mixed-reality environments. We show that even when the time required to complete some tasks is longer when using Myo-supported gestures, users often preferred to use them as they are highly intuitive for 3D object manipulation. As part of our design considerations we have discarded the use of sustained gestures identified as uncomfortable in the literature, e.g., wrist adduction or the extension called “wave-out” by Myo. However, our user study suggests that one should include these as well in a system, indicating users’ preference to have available a larger set of gestures for more flexibility during interaction.

Concerning limitations for EMG-based interfaces, we conclude that a higher resolution is extremely important to support accuracy and robustness among a large set of users. Challenges of this technology includes the need of a better calibration process for each user since the observed variance among users was significant. Ergonomic issues must also be addressed.
and becomes especially relevant when carrying out tasks for longer times. In addition, we suggest the further study of implantable EMG solutions as its technology advances with adequate safety standards and becomes more accessible for the general public.

REFERENCES


