A Roadmap to Prolific Embodied Synthesis: A Narrative Review of Motion Capture Systems for Music Generation

A Roadmap to Prolific Embodied Synthesis

A qualitative analysis of user experience and a quantitative comparison of mocap hardware performance and cost

Landon L Odishaw-Dyck

Mount Royal University, Calgary, Canada, lodis898@mtroyal.ca

The creation of music via the capture of body motion, or embodied synthesis, has suffered from a lack of purpose-built hardware. To aid in the proliferation of embodied synthesis, we compare the performance and cost of multiple motion capture systems. It was found that many systems suffer from variability that would greatly diminish the virtuosity of embodied synthesis. Furthermore, most systems are inaccessibly expensive for most musicians. Fortunately, there is potential for future developments to provide consistent performance and at an accessible price point via the creation of hybrid hardware systems. We also investigated the user's experience of each system. We posit the notion that the use of human joint angles provides the most intuitive form of musical generation, although each type of hardware provides unique experiences, thus underlining the potential of hybrid systems. Finally, we outline the future development of a system with sufficiently low error, latency, and cost.

CCS CONCEPTS • Human-centered computing~Human computer interaction (HCI)~Interaction techniques~Gestural input • Human-centered computing~Human computer interaction (HCI)~HCI design and evaluation methods~User studies • Applied computing~Arts and humanities~Performing arts • Applied computing~Arts and humanities~Media arts • Applied computing~Arts and humanities~Sound and music computing • Human-centered computing~Human computer interaction (HCI)~Interaction devices~Haptic devices

Additional Keywords and Phrases: Motion capture, Embodied synthesis, Embodied interaction, Music synthesis, Digital instruments, Virtual instruments, Tangible user interfaces, Interactive music systems, Virtual reality, Augmented reality, Extended reality

1 INTRODUCTION

The article explores the suitability of motion capture (mocap) systems for the creation of music via the mapping of human motion to synthesizing software. This process can be referred to as embodied synthesis. In [1], Turchet et al. surveyed studies that explored the synthesis of music using extended reality hardware, such as VR headsets and controllers. Many attempts suffered from a lack of specialized input hardware, emphasizing the need for development of purpose-built systems to enable embodied synthesis to rival the nuanced, precise, and tactile interfaces offered by traditional instruments. This article takes a holistic view of embodied synthesis, from the performance and characteristics of each type of mocap hardware, to how these factors affect user experience. This article is structured as follows: first, we set latency and spatial accuracy targets to ensure a virtuous musical instrument; then relevant methodologies are introduced, followed by an explanation of the operating mechanisms of each type of mocap hardware; from there we analyze the user's experience of three common mocap systems; finally, we analyze the cost, performance, and weaknesses of each system, along with an exploration of hybrid systems.

1.1 Virtuosity and Difficulty

In [2], Wessel et al. argued the importance of balancing the initial difficulty of learning a new digital instrument with its potential for a high virtuosity ceiling. The difficulty should be manageable such that it does not dissuade new players. Simultaneously, the instrument should avoid being so simple that it limits the potential for intimate control such that experienced players cannot continually progress their skills and expressivity. What level of accuracy and latency is required to provide a high virtuosity ceiling?

1.2 Latency Requirements

Although many real-time applications can withstand up to 45 milliseconds of latency [3], musical synthesis has stricter standards. In [2], Wessel argues that digital instruments should have a maximum latency of 10 milliseconds between motion and sound, for the sake of rapid feedback. Furthermore, to maintain repeatability, the standard deviation of latency should not exceed 1 millisecond [2]. Latency deviation will become more obvious when multiple users attempt to synchronize their play. Furthermore, increasing user counts will increase computational workloads, underlining the importance of minimizing latency. Note that, although additional computing power can lower latency [4], it will also increase costs. To leave time for synthesizing software, hardware signal processing and transmission must be under 10 milliseconds.

1.3 Spatial Requirements

The output of physical instruments will consistently conform to any subtle changes of the user's input [5]. Therefore, a high level of spatial performance is required for embodied synthesis to mirror this behavior. In [6], Filippeschi et al. argued that consistency is crucial for control purposes. Specifically, it was contested that given minimal spatial variance in the mappings that control the output, a human can adapt their motion to compensate for any inaccuracy. There is minimal research on the spatial requirements of embodied synthesis. It is common in the literature to validate mocap systems by comparing their performance with that of optical mocap [7], which has been the gold standard of mocap for decades. Joint angles and position are the most common metrics of assessing mocap accuracy and will thus be used to compare hardware. Spatial performance is dually as important when a user has limited mobility, range of motion, or dexterity, as can be highlighted by [8]. Therefore, spatial accuracy can be seen as having a direct impact on the accessibility of the system [1], allowing those who are unable to play conventional instruments to leverage their given mobility as a control mechanism. Note that spatial inaccuracy is exacerbated when multiple users directly interact with each other.

1.4 Cost Requirements

Finally, the proliferation of embodied synthesis largely depends on the monetary cost of the hardware [1]. By lowering prices, such that they are comparable to traditional acoustic instruments, we increase accessibility. Thus, allowing for more user feedback which can then be used to further improve the system and further assist in its proliferation.

2 METHODOLOGIES AND HARDWARE

With our performance standards set, we introduce relevant methodologies. Different mocap hardware has unique characteristics. To understand these eccentricities, we explain relevant terminology and the operating mechanisms of each type of system, including the processes by which each system obtains different metrics of human motion.

2.1 Terminology

Throughout the article, the terms relative, absolute, and translation will be used. A relative reading is a measurement of a given body segment that is defined in relation to other body segments, and it is affected only by changes in joint angles. Relative measurements are inherently isolated, in that motion from distal segments do not affect measurements for a given segment. An absolute reading is a measurement that is defined within an absolute coordinate system, with axes situated in the surrounding capture space, external to the user. Absolute coordinate systems may be global, providing a common coordinate system for multiple users. Absolute measurements are inherently permeable, in that motion from body segments that are distal to a given segment can affect measurements of that given segment. A translation is a change in the absolute position of the user, without any change in relative position. As such, this motion cannot be tracked by relative measurements, only by absolute measurements. For example, jumping is only completely trackable via an absolute system. Relative measurements would only reflect joint adjustments needed to perform the jump but would not relay the displacement of the jump itself.

2.2 Kinematics

Kinematic models contain knowledge that define the relationships amongst kinematics chains. The term kinematic chain describes a series of rigid body parts and the joints that connect them. For example, the leg is defined by the length of two rigid bodies, the hip and calf, and the hinge joint that connects them. There are two relevant methods of applying kinematic models. Forward kinematics is the calculation of the position of the end of a kinematic chain given the joint angles. Conversely, inverse kinematics is the determination of joint angles given the chain's end position. Generally, inverse kinematics is more computationally expensive, adding more latency than forward kinematics [9]. Although exact estimates of the latency are beyond the scope of this article, applying inverse kinematics to optical position data can take tens of milliseconds [10]. Note that raw hardware measurements may not relay all relevant kinematic information. For example, the relative distance between the chest and a forearm does not convey whether the forearm is in front or behind the chest. This type of kinematic information could be vital in properly conveying the user's desired musical outcome. This underlines the importance of applying kinematic models. As we will see, kinematics is used differently by different mocap hardware.

2.3 Operating Mechanisms and Metric Calculations

We can categorize mocap systems by the metrics that they most directly sample [11]. If a system can directly acquire a metric, it implies that it can obtain a measurement with minimal processing, and thus with minimal latency and potential error. Now we will explain how each relevant metric is either directly sourced or calculated for the three hardware systems we are primarily interested in.

2.3.1 Optical Systems

Optical motion capture uses the 2D images from two or more cameras to reconstruct a 3D position of the light-reflecting markers placed on a user. Commercial systems use approximately 30-40 markers to model the whole body [12]. Optical systems directly sample the absolute global position of markers. Then, absolute linear velocity and acceleration is acquired indirectly via mathematical derivation of absolute position. From absolute position we also calculate relative position, which in turn allows the mathematical derivation of relative linear velocity and acceleration. To calculate absolute orientation requires the absolute position of at least three noncollinear markers placed on a given rigid body. Given absolute orientation, we derive absolute rotational velocity and acceleration.

From absolute orientation, we also calculate relative orientation, which in turn allows us to derive the relative rotational velocity and acceleration. Finally, applying inverse kinematics to absolute position and orientation gives us joint angles, which in turn allows joint-angular velocity and acceleration to be derived.

2.3.2 **Inertial Systems**

Inertial motion capture utilizes inertial measurement units (IMU), which consist of micro-electromechanical (MEMS) gyroscopes, accelerometers, and magnetometers. The actual process of sampling IMU data involves fusing readings of all three sensors together. [6] provides details. By attaching IMUs to each rigid body of the user, we can reconstruct a full body. Most commercially available IMU suits provide between 15 and 40 IMUs [12]. Each IMU directly provides absolute linear and rotational acceleration within its own absolute coordinate system. Via mathematical integration, we then acquire absolute linear and rotational velocity, with further integration yielding absolute position and orientation for each IMU. To position multiple IMUs on a given user within a common absolute coordinate system, a calibration pose and kinematics are used. To orient the IMUs within a common absolute coordinate system, each IMU uses gravity and the Earth’s magnetic field. From absolute position and orientation, we calculate relative position and orientation. From relative position, we derive relative linear velocity and acceleration. From relative orientation we derive relative rotational velocity and acceleration. Finally, using absolute position and orientation, and inverse kinematics we acquire joint angles, which in turn allows joint-angular velocity and acceleration to be derived. Although there are multiple approaches for acquiring each of these measurements [6] the preceding paragraph has outlined the basic process to understand how direct each reading is. Note that IMU suits track their absolute position in relation to their starting position but cannot provide a global coordinate system common to multiple users.

2.3.3 **Mechanical Systems**

Classical mechanical systems attach to the user an exoskeleton which contains linear and angular electromagnetic potentiometers, that mirror the behavior of the body's joints. Up to 42 potentiometers are required for full body capture [13]. Mechanical systems directly acquire joint angles. From joint angles we derive joint-angular velocity and acceleration. Using forward kinematics, we obtain relative position and orientation. Then we derive relative linear velocity and acceleration, as well as relative rotational velocity and acceleration. Unfortunately, purely mechanical systems cannot acquire any absolute readings.

2.4 Soft Tissue Artifacts

Note that all motion capture systems are vulnerable to soft tissue artifacts. Human motion is primarily defined by the pose of the skeleton, while all hardware is placed on the skin. Soft tissues can move separately from the skeleton, such as the expansion of muscles, resulting in an error of pose estimation [14].

2.5 Hardware Placement

Finally, notice that IMUs can only directly provide absolute linear acceleration for their own location on a body segment. To obtain the acceleration of a point adjacent to the IMU requires integration, the application of kinematics, and finally derivation. Similarly, optical systems require kinematics to obtain the absolute position of a point adjacent to a marker. Thus, to acquire these measurements directly, IMUs or markers should be placed at all locations on the body that are used for a control mechanism. These locations are difficult to predict due to the vast creative possibilities of human motion and will require user feedback [1].

3 USER EXPERIENCE OF METRICS

In summary, each system can directly sample a different metric. IMUs can directly acquire absolute linear and rotational acceleration. Optical systems directly acquire absolute global position. Mechanical systems directly acquire joint angles. To minimize latency and error, future development should be centered on hardware that offers direct measurements of the most intuitive metrics. So which metric provides the best user experience?

3.1 Imparting Energy

Humans intuitively understand how acoustic instruments respond to our physical input of energy, such as a strum, pluck, or strike [5]. Hunt et al. found that users became engrossed with digital instruments that mimicked acoustic instruments by requiring continuous motion to produce sound [15]. As such, it seemed that their own energy was imparted to the instrument. Indeed, when using a mechanical suit to synthesize, Collins et al. found a user preference for control mechanisms that required a sudden motion [16]. To leverage this notion of imparting energy, we can use velocity and acceleration. Velocity can either be derived from an optical or mechanical system or integrated from IMU acceleration. When investigating embodied synthesis, Skogstad et al. noted that users preferred IMU based velocity and acceleration, while velocity and acceleration derived from optical data suffered from noise and required smoothing [12]. Therefore, IMUs seem best poised to directly facilitate the imparting of energy.

3.2 Permeability

Of course, the absolute readings of IMUs are permeated by motion from distal body segments. To address this permeability, Skogstad et al. used pattern recognition to specify conditions for the activation of control mechanism [17]. For example, a user’s hands had to be close together to enable the triggering of a clap, otherwise their acceleration was ignored. Yet, a single recognition failure can cause a major disruption to synthesis [18]. Given our need for extremely high repeatability for the sake of music synthesis, using absolute measurements for control remains challenging.

3.3 Proprioception

Research on proprioception can provide further guidance. Proprioception refers to the body's ability to sense the position and motion of its various segments without relying on sight. Darling et al. found that healthy, blind-folded adults can consistently determine the location of their hands in relation to each other with under 20 millimeters of mean error and under 8 millimeters of standard deviation [19]. Meanwhile, Barden et al. got blind-folded participants to repeatedly point to the same absolute position in a 3D coordinate system. Participant error ranged from 60 to 100 millimeters [20]. These results indicate that humans consider relative movement more intuitive than absolute movement. A mechanical system is the only system that directly provides data in a relative coordinate system. Note, relative measurements cannot register translations. Failing to track a translation may break the user’s perception that energy is being imparted.

Many applications, such as film production, aim to capture motion without the user being aware of the hardware. Conversely, an awareness of the hardware’s operations is beneficial for control applications, specifically in embodied synthesis, where users are deprived of the haptic feedback of traditional instruments. Khabie et al. found that wrapping an elastic bandage around a participant’s elbow improved their proprioceptive accuracy [21]. The innate friction of mechanical systems can provide passive haptic feedback in a similar manner. Furthermore, this passive haptic feedback may serve the perception that energy is being imparted into the instrument. Going even further, many hand motion capture systems provide active haptic feedback [22]. Mikula showed that providing vibrotactile stimuli can improve proprioception, although the hardware needed to provide active feedback increases cost and weight [23].

3.4 Separated and Integrated Parameter Control

Human-computer interaction research further supports the use of mechanical systems. Note that digital instruments can separate musical parameters, such as amplitude and frequency, that are usually integrated for acoustic instruments [24]. Jacob et al. argues that each parameter of the task that is perceivably distinct should have a separate input parameter within the control mechanism [26]. Supporting this notion, Collins et al. found that mapping separate roles for each limb provided a greater sense of control when using a mechanical suit to synthesize [16]. Alternatively, parameters can be combined and controlled by a single integrated mechanism [15]. Rovan et al. noted that while beginners prefer separated mechanisms, experienced musicians prefer integrated mechanisms [24]. Therefore, allowing for both integrated and separated control facilitates both a low barrier to entry and a high skill ceiling.

Wagner et al. found that healthy humans are adept at moving a chosen joint while keeping surrounding joints still [25]. Meanwhile Jacob et al. noted the difficulty of moving straight along a single vector in a 3D coordinate system, thus confounding the task of isolating parameters using directionality [26]. All these points together validate the use of joint angles as opposed to 3D vectors, to provide both separated and integrated control, complementing different levels of user ability.

3.5 Direct Collaboration

Despite these strengths of relative measurements, global absolute coordinates provide certain unique opportunities. Turchet et al. advocated that mocap be fully leveraged to offer musical interactions that are not possible in the real world [1]. For instance, mocap is uniquely well suited to facilitate collaboration, such as the shared control of the same instrument. Collins et al. noted an interest in transforming one user's input via the motion of another user [16]. Optical systems provide a global capture space, thus allowing users to collaborate in a common coordinate system.

3.6 User Experience Summary

Let us summarize. The use of acceleration facilitates the perception of imparting energy. Relative motion is more intuitive than absolute motion. Passive haptic feedback can improve proprioception. Using joint angles allows for both separated and integrated control mechanisms. A common coordinate system enables direct collaboration. If we were relegated to a single system, the literature indicates that mechanical suits would provide the most intuitive metric. Using only one system and its most direct metric provides very low latency variation, as the calculation of indirect metrics results in variable latencies between metrics. Yet, each system provides a distinct experience for the user, thus it is worth exploring hybrid systems, allowing for the direct sampling of multiple metrics. Now we will evaluate the performance and cost of each system, while simultaneously exploring the ability of constituent systems to compensate for each other's weaknesses.

4 OPTICAL SYSTEM ANALYSIS

Topley et al. found that most commercial optical systems have around 0.2 millimeters of positional error on average, with a standard deviation of roughly 0.1 millimeters [27]. For angle measurements, they have around 0.2 degrees of error on average, with a standard deviation of approximately 0.1 degrees. The primary weakness of optical systems is their vulnerability to occlusions, which is when a marker is not visible by at least 2 cameras, and thus we lose readings for that marker. There are methods to provide estimates of the position of occluded markers using the trajectory of recently visible motion [28]. However, using estimates undermines control and virtuosity, as the subtleties of user input become less consequential. Increasing the number of cameras can minimize the frequency and duration of occlusions, but this increases costs significantly. Regardless of the number of cameras, occlusions are inevitable, and will cause crippling interruptions to the synthesis process. Unfortunately, all occlusion resistant alternatives to optical systems have major caveats in performance.

4.1 Occlusion Resistant Alternatives

Magnetic systems are limited in the number of users that can be tracked. Additionally, they are vulnerable to magnetic distortion, which can lead to inconsistent readings [29]. Additionally, they have cumbersome wired setups [12]. Ultra-wideband systems are vulnerable to multipath signal propagation, which is the scattering of the signal waves from the occluding objects, providing inconsistent readings [30]. Zhao et al. used ultra-wideband technology with error on the scale of hundreds of millimeters [31]. Ultrasonic systems are also vulnerable to multipath signal propagation [32].

Inside-out tracking is an emerging computer vision method, which uses multiple outside-facing ultra-wide-angle cameras placed on a given segment of the user. This method captures and maps multiple high-contrast objects in the capture space, such as the corner of a window frame. Then, using a computer vision technique called visual-inertial simultaneous localization and mapping, or SLAM, creates a reconstruction of the room to locate itself in the space [33]. SLAM is usually used with the assistance of IMUs. But unlike purely IMU mocap, SLAM can provide global position readings. SLAM is commonly used for virtual reality headsets, although smaller models are emerging that can be placed on other parts of the body, thus allowing full-body tracking. Monica et al. evaluated the inside-out tracking accuracy of the Oculus VR headset and found an average error spatial error of 18.3 millimeters, weaker than traditional optical methods. Furthermore, SLAM is reliant on its surroundings, which may be dynamic or unfavorable, such as when it is too dark [34]. Finally, each inside-out tracking node requires significant CPU power, increasing costs. Although they have potential, SLAM methodologies are currently too inaccurate and variable in comparison with traditional optical mocap.

Evidently, all optical alternatives that provide global absolute data are susceptible to some form of interference. Despite occlusion vulnerability, optical systems can guarantee the consistency of their given readings. Before discussing the incorporation of IMUs and mechanical suits to resolve occlusions, the challenges of marker identification, cost, and latency will be addressed.

4.2 Passive Markers

The most common form of optical mocap uses retroreflective markers, which reflect the emissions of infrared lights attached to each infrared camera. All markers are indistinguishable from one another. Thus, each marker must be mapped to a body segment to digitally reconstruct a user's body. After initialization, multiple methods are used to maintain the distinction between markers and to re-identify recently occluded markers. One of these methods uses their previous positions and apparent motion to predict the trajectory of each marker [28]. If these methods fail, then markers will be misidentified, and this must be fixed in post-production, thus severely limiting real-time usage.

4.3 Active Markers

Using active markers addresses the issue of misidentification, as each marker identifies itself using an LED. Furthermore, the range of active markers is greater than that of passive systems [11], lowering the required number of cameras. There are multiple methods of active marker identification. The simplest method activates a single marker for one frame. This means each additional marker requires its own frame, lowering effective frame rates and increasing the latency for a full body capture [35].

4.4 Temporal Patterns

Alternatively, we can identify using unique sequences across successive frames. To encode these unique identifying sequences, we can use brightness or wavelength. The simplest approach is to create a binary sequence, with each marker being on or off. Alternatively, we can use differing levels of brightness to encode with a higher base, depending on the brightness resolution of the camera. We can also create encoding sequences using wavelengths.

The Nyquist-Shannon Theorem states that an analog signal must be sampled at a rate of at least twice the highest frequency present in the signal. In other words, we need at least two frames for each distinct value in the identification sequence. As the number of required IDs increases, so too shall the length of the identification sequence. In the intervening frames, while the identification process takes place, we must rely on passive tracking methods such as trajectory estimation [28]. This increases the chances of marker misidentification. Additionally, temporal sequences require precise synchronization amongst all cameras and markers. The only manner of circumventing these limitations is to use unchanging wavelength or brightness as a means of identifying markers, rather than a temporal sequence. The issue with this approach is that we are limited in the number of unique IDs as dictated by the wavelength or brightness resolution of the cameras. Finally, commercial systems predominantly use monochrome CMOS sensors, which do not measure wavelength, only light intensity [36], thus limiting the use of wavelength encoding.

4.5 Markerless Methods

Advances in computer vision have enabled markerless systems that require less setup time. Unfortunately, these methods suffer from variability due to the dynamic nature of shifting clothing, inconsistent lighting, and the contrast between users and their backgrounds [37]. As such, computer-vision based approaches have yet to achieve the sub-millimeter accuracy [38] which traditional optical systems are capable of. Furthermore, the computational complexity of the post-processing required compared to a marker-based approach results in additional latency [39].

4.6 Price of Optical Systems

The cost of commercial optical mocap is prohibitively high for most parties except larger institutions and companies. This is clear from figures 1 and 2, which plot the number of required cameras, and their price, per user in the capture space, as recommended by the OptiTrack website. OptiTrack is one of the more affordable commercial optical solutions. Note, these costs are for the cameras themselves, not including other hardware such as cables and hubs, nor the cost of subscriptions.

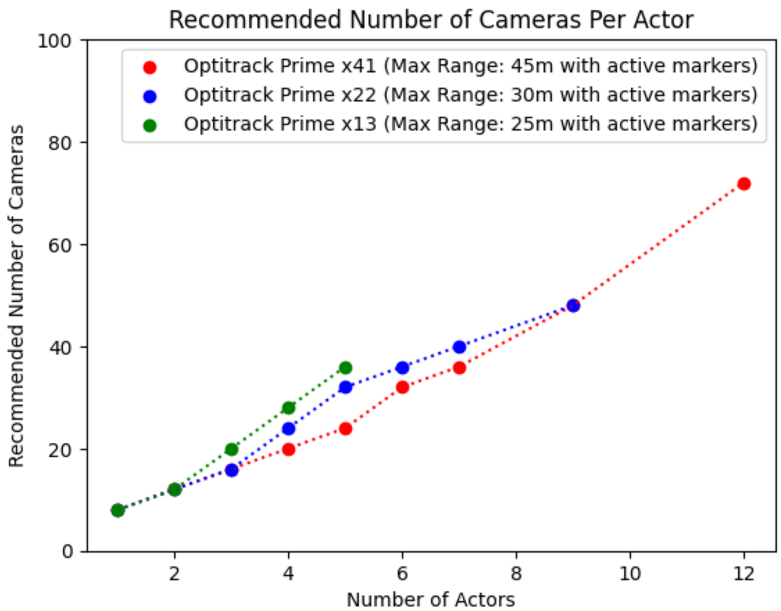

Figure 1: A plot of the number of users in a capture space versus the number of cameras needed to capture those users, via OptiTrack (https://optitrack.com/systems/)

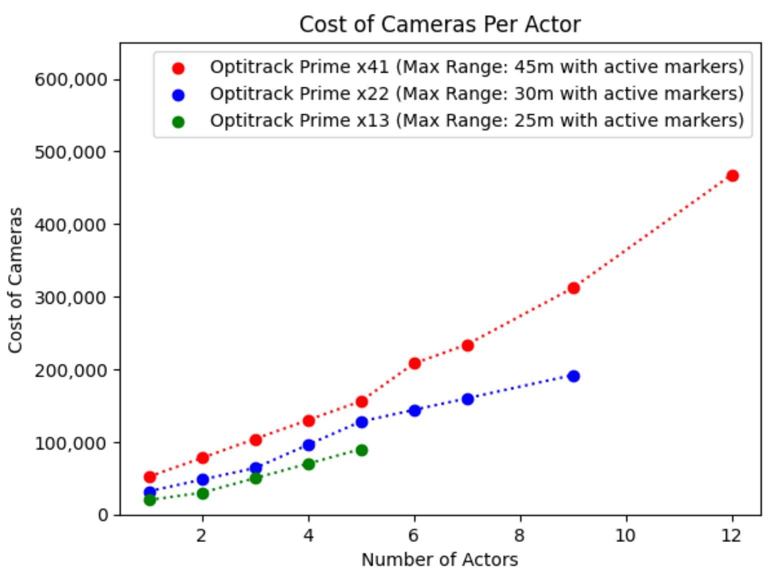

Figure 1: A plot of the number of users in a capture space versus the total price of the cameras needed to capture those users, via OptiTrack (https://optitrack.com/systems/)

Table 1: OptiTrack Camera Model Specifications and Mean Camera Count and Cost Per User

OptiTrack Camera Model	Model Cost (USD 2023)	Resolution	Frames Per Second	Mean Numbers of Cameras Per User	Mean Camera Cost Per User
Prime x13	$2,500	1280 x 1024	240	6.97	$17,433
Prime x22	$4,000	2048 x 1088	360	6.10	$24,390
Prime x41	$6,500	2048 x 2048	180	5.66	$36,792

4.7 Smartphone Potential

The proliferation of high-quality cameras in smartphones offers an affordable alternative to commercial systems. In Figures 3 and 4, we graphed the frame rates achieved at various resolutions for the flagship smartphones of Apple and Samsung. As of 2023, these two companies collectively account for over half of the global smartphone market share [42].

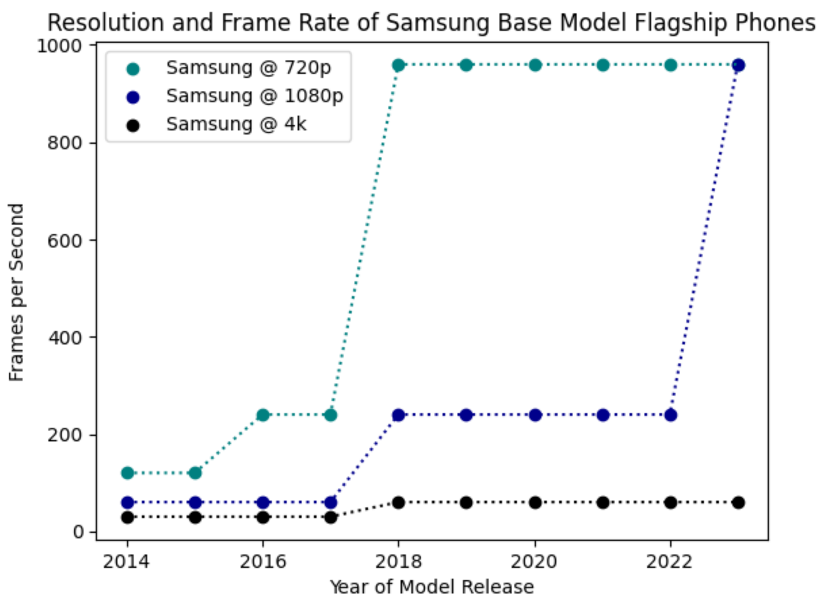

Figure 3: A plot of the Samsung's flagship smartphone from 2014 to 2023 versus the maximum frame rate they can achieve at various resolutions, via GSM Arena (https://www.gsmarena.com/)

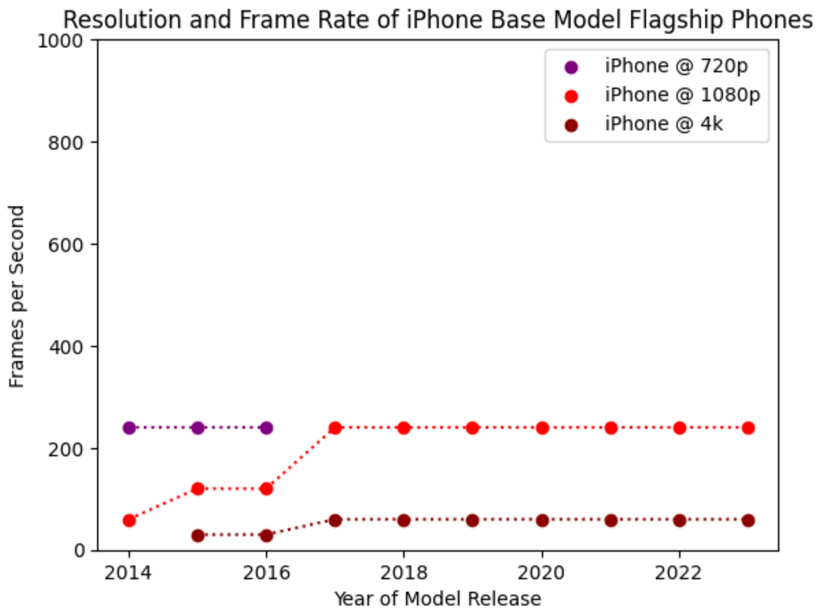

Figure 4: A plot of the Apple's base model flagship iPhone from 2014 to 2023 versus the maximum frame rate they can achieve at various resolutions, via GSM Arena (https://www.gsmarena.com/)

We must strike a balance between resolution and frame rate. It is recommended not to use any resolution lower than 720p [41]. Although 4K has been available at 60fps from both companies since 2017, this is close to the lowest advisable frame rate for real-time interaction [42]. Thus, the best balance between resolution and frame rate seems to be 1080p at 240 fps. Resolution has a direct impact on detection range and thus the volume of the capture space. Using only ambient room lighting, neon-tape, and a single iPhone recording at 1080p Vincent et al. could discriminate markers at distances up to 18.5 feet [37].

4.8 Frame Rate Synchronization

Interoperability between smartphone models, including between companies, will be crucial for accessibility and proliferation of a phone-based mocap. Frame rate synchronization is of primary concern. Optimally, all camera shutters would be synchronized with each other. Unfortunately, actual frame rates can differ slightly from the manufacturer specifications. Teramoto et al. found that their iPhones had a frame rate of 60.001 FPS [43]. As such, they recommended testing each phone before usage. To address the accumulated offset of shutter timings we can resynchronize cameras, although this may lead to a noticeable stutter. Teramoto et al. opted instead to interpolate between frames instead of synchronizing two cameras with staggered shutter timings. One camera would interpolate the position of a given marker between two of its frames to match the time of the frame captured by the other camera. Of course, there is error associated with this interpolation. Further work must be done to address this deviation from manufacturer specifications.

4.9 Wavelength Capabilities

Smartphones use RGB CMOS sensors, allowing for the use of wavelength as an identification method. RGB CMOS sensors have wavelength resolutions down to 1 nanometer [44]. Unfortunately, smartphone cameras use filters to block infrared and ultraviolet light [45], thus limiting us to the visible spectrum, which is approximately 380 - 750 nanometers.

4.10 Additional Smartphone Capabilities

Smartphone components other than the camera can facilitate mocap. Commercial mocap cameras have specialized onboard processing units to extract markers from a captured image [46]. Given their inherent computing power, smartphones are equipped to handle this processing.

Additionally, the other sensors of a phone can contribute. As occlusions change over time, the set of cameras that are used to reconstruct the position of a given marker changes as well. Skogstad et al. noted that this transition between camera sets over time can lead to positional error [12]. This is due to slight calibration errors during camera setup. Thus, as setup quality increases, camera set changes become less noticeable. Meanwhile, Skogstad noted that the stress of resolving technical difficulties can be detrimental to the creative process. Therefore, we must aim for quick yet effective setup methods. To these ends, Teramoto et al. leveraged the onboard flashlights and microphones of two smartphones to develop a calibration technique [43]. First, the flashlights were used to synchronize the two phones. Then, using time-of-flight methods, the microphones were used to triangulate the location of the phones in relation to one another.

4.11 Optical Latency

Finally, we estimate the latency of commercial systems. To start, scanning the camera's sensor takes between 2.8 and 10 milliseconds, depending on the camera model [47]. Then the camera's onboard processor distinguishes the light of markers from external light pollution that should be ignored. This process is called point of interest extraction (POI). Then the center, or centroid, of each marker is determined. These two steps take approximately 0.5 milliseconds for one user. Active markers relay their own centroid [46], thus eliminating some of this delay. From here the data is transmitted to a PC via ethernet cables, taking only microseconds. Then the PC reconstructs data from the cameras, taking 1 to 5 milliseconds for one user. Thus, theoretically, one user's position can be captured in 4.3 milliseconds using passive methods [35]. To do so will require meticulous management of the processing pipeline. Suboptimal conditions, such as an overburdened PC, can result in up to 25 milliseconds of latency [48]. Maintaining low latency with smartphones will be an important challenge. It will likely be necessary to use a bare-metal programming approach. Note that optical position data requires filtering before its use as a control mechanism, adding latency. Refer to [5] for a comprehensive introduction to the topic.

Now, with the issues of marker identification, cost, and latency addressed, we can discuss the incorporation of IMUs and mechanical suits to compensate for occlusions. Although position from IMUs and mechanical systems is indirectly sourced, the inconsistencies of occlusion-resistant optical alternatives leave us no choice.

5 INERTIAL SYSTEM ANALYSIS

Now, we will discuss the advantages of an inertial-optical hybrid, the challenges of IMU based position, IMU latency and cost, followed by a discussion of future research.

5.1 Symbiotic Inertial-Optical System

While IMU suits can measure position changes in relation to their origin, optical assistance is needed to locate users in relation to each other. Furthermore, purely IMU systems struggle with elevation changes, such as stairs, and thus often require post-processing [12]. Additionally, optical assistance negates an IMU's reliance on Earth's fickle magnetic field to orient itself. Note, a hybrid system can run momentarily on only IMU readings. Therefore, by combining a 100 FPS optical system and a 1000 FPS IMU system, we get frame rates comparable to optical systems of exorbitant cost [49]. Most importantly, optical assistance can constrain the long-term drift of IMU-based position readings [49].

5.2 Inertial Drift and Accuracy

IMU position data is primarily sourced via the doubly integrated acceleration readings of the accelerometer and gyroscope. Both the raw readings and the integration process yield errors that accumulate over time, thus leading to positional drift. There are many approaches to increase accuracy and constrain drift within purely inertial systems.

5.3 Assumptive Methods

Many attempts hinge on assumptions or restrictions regarding the motion of the user. For example, Zandbergen et al. improved accuracy by relying on the assumption that during running, the leg follows a cyclical motion [50]. Taking a less assumptive approach, Yang et al. trained a neural network to recognize a given set of motions. Although this reduced error for the given set of motions, it did not generalize well to other motions [51]. Another common approach is to use the sudden deceleration of foot-ground contact to constrain drift [52]. Although worth noting, it does not generalize to other body parts that do not consistently contact external objects. Despite their efficacy for specific motions, these methods generalize poorly, and may be unresponsive to dynamic and improvisational synthesis. Therefore, we must explore less assumptive methods.

5.4 Kalman Filtering and Kinematics

A common and generalizable method to improve results is to employ filtering algorithms to fuse readings from the constituent IMU components. This filtering mitigates the noise and bias of the raw readings from individual IMU components. There are many filtering methods [6], but the Kalman filter is the most prolific. In short, a Kalman filter iteratively refines motion estimates by using both new and previous readings along with their corresponding uncertainties. By using different types of Kalman filters, we can improve our trajectory estimates, but at the cost of greater computational complexity and thus latency. For example, the standard Kalman filter assumes linear motion, thus requiring fewer computations, but obtaining less accuracy than a more complex approach. Extended and unscented Kalman filters model non-linear motion, though at the cost of increasing computational complexity, sometimes exponentially [53]. Another generalizable approach is the use of kinematics to constrain errors. For example, a knee joint only has one degree of freedom, and thus any motion in a different direction can be considered an erroneous reading [6].

5.5 Long-Term Drift

Skogstad et al. and Li et al. reported drift rates for purely IMU mocap suits from Xsens and Perception Neuron respectively. Both systems applied Kalman filters and kinematics [4, 49]. For constantly moving users, Skogstad et al. noted an accumulating error of up to 22.2 millimeters per second, whereas Li et al. recorded a cumulative drift rate of 0.78 millimeters per second. Skogstad et al. noted a drift rate of 2.5 millimeters per second for stationary IMUs. Thus, kinematics and filtering alone are unable to constrain drift. By combining an optical system and an IMU system, Li et al. showed that cumulative drift can be resolved by optical assistance [49]. Despite this, the mean error between the optical and IMU position readings was just over 10 millimeters, indicating discrepancies between the systems.

5.6 Occlusion Performance

With long-term drift addressed, we will explore how effectively IMUs can emulate optical position during occlusion. Because commercial systems maintain their methods as trade secrets [54], obtaining data on hybrid occlusion performance has proven difficult. To obtain insight, we must get creative with the publicly available research.

5.7 Linear Kalman without Kinematics

In [6], Filippeschi et al. investigated the positional error of IMUs when using a Kalman-based method proposed by Yun et al. in [55], but without kinematics applied. Specifically, Filippeschi et al. observed kinematic chains of limited length, namely motions involving the elbow and shoulder joints, while the rest of the body was kept still. This mirrors the performance of a hybrid system where only small segments of the body will rely solely on IMUs. The first trial involved repeatedly flexing, then extending only the forearm, and thus only one IMU was used. The second and third trials involved shoulder extension and abduction respectively, thus including two IMUs and a longer kinematic chain. All three trials used low acceleration motions. The first trial indicated an average error of 89.2 millimeters and a correlation value of 0.77 for the forearm IMU. Unfortunately, the second and third trials yielded error averages of 121.4 and 243.8 millimeters respectively, and correlation values of 0.86 and 0.36 respectively. Thus, longer kinematic chains yield greater error.

5.8 Non-linear Kalman with Kinematics

By applying a more computationally complex approach, we can improve results. Atrsaei et al. used kinematic constraints and an unscented Kalman filter which is designed to estimate non-linear motion [56]. Like Filippeschi et al., Atrsaei et al. investigated shoulder motion with two IMUs, except the shoulder extension and abduction were combined. A low-acceleration trial yielded an RMSE of 38 millimeters for the forearm IMU. From this we can see that nonlinear Kalman filters and kinematics can provide greater accuracy than linear Kalman filters, but at the cost of greater computational complexity and therefore latency [53]. In addition to this worrying RMSE, Atrsaei's data showed regular error fluctuations approaching 100 millimeters. Note that Atrsaei performed this same trial, but with greater acceleration. This resulted in 44 millimeters of RMSE. Therefore, increases in acceleration, and the resulting increase in error, increases the variability of any control mechanism.

5.9 Joint Angle Performance

The joint angle results are similarly inaccurate. Sfalcin et al. showed that Xsens suits have an RMSE of 24.5 degrees, with errors as high as 76.9 degrees [57]. The Perception Neuron suit had an RMSE of 35.2 degrees, with a maximum error of 93.2 degrees.

5.10 Inertial Latency

Now to characterize latency. Xsens offers a wired system, where all IMUs are wired to a central on-person unit, which is then wired to the PC. This removes the latency and variability associated with wireless connections [48], offering a latency estimate of full-body capture, excluding hands. Using onboard filtering and integration, and the application of kinematics, Xsens achieves 20 milliseconds of latency [58], double our target.

5.11 Inertial Cost

A suit from Xsens, with 17 IMUs, costs $4,600.00 (USD 2023) for one user. Individual IMUs can be purchased for $325.00. Each IMU has specialized onboard processing power to filter mathematically integrate the data. This cost excludes hand-tracking hardware and software subscriptions. IMUs that do not have onboard processing power for integration or filtering are much cheaper, commonly costing under $10.00 per unit.

5.12 Hybrid Inference

Before drawing final conclusions regarding inertial-optical hybrids, we will briefly describe some potential improvements that are largely unexplored within the publicly available literature. Li et al. combined turn-key solutions, where each system completed motion estimates, after which their respective estimates were merged [49]. This approach does not leverage the unique hybrid methods of acquiring measurements, such as fusing multiple lower-quality sources to improve estimates and limit drift, like the Kalman filtering of IMU components. For example, optical position, orientation, and kinematics, can be fused with IMU readings from an adjacent occluded limb to better infer its position. This can be done iteratively down a longer kinematic chain.

5.13 Minimizing Hybrid Latency

Hybrid systems offer many methods to acquire data, but each method requires resources. Santos et al. found that a hybrid using a turn-key solution created excessive computational demands and latency over our target [48]. Therefore, we must avoid the computation of higher error readings when a better one is available. For example, integrating IMU data to obtain position or using adjacent optical data to infer position is only necessary during occlusion. To cease unnecessary computations requires the tracking of error as it propagates from initial hardware sampling to completion of computations, including mathematical integration and derivation, and the application of kinematics. This entire process may be distributed across IMUs and the central PC. Thus, a comprehensive and efficient solution that utilizes hybrid inference and minimizes latency requires a custom ground-up application with full access to the entire processing pipeline of both systems, including error estimates. Due to the closed-source software of industry leaders, the creation of such a comprehensive hybrid system using turn-key solutions becomes severely complicated [59, 6]. Without conducting experimentation of these unique hybrid properties, any final conclusion on the latency and accuracy of inertial-optical hybrids is incomplete. Such exploration will require a full optical and IMU hardware setup, which is beyond the scope of this article. Regardless, we have gained some critical insights.

Evidently, IMUs struggle to provide a sufficient balance between high accuracy and low latency, thus limiting their ability to emulate optical position. While increasing the complexity of the filtering method can increase accuracy, this increases latency. To address this, we could increase computational power, although this increases hardware costs which lowers accessibility. Thus, it seems only the most direct readings of IMUs, i.e. linear and rotational acceleration, can be acquired with sufficiently low error, cost, and latency.

6 MECHANICAL SYSTEM ANALYSIS

Mechanical systems are the only alternative left to deal with occlusions. From here determine their latency, cost, and accuracy. Then we explore how mechanical data can be inserted into an optical system. We finish by discussing mechanical ergonomics.

6.1 Mechanical Latency

Sim et al. processed and filtered the joint angles of a single finger in 0.15 milliseconds, from two angular potentiometers, as opposed to the 1.92 milliseconds required to apply a Kalman filter to a single IMU [60]. Based on their results, modeling all joints, excluding the hands, a mechanical system will take approximately 3 milliseconds. This is without wireless transmission, nor the calculation of other metrics. Sim et al. fully processed the hand data of ten angular potentiometers in 4 milliseconds, including wireless transmission via Bluetooth [60]. Fortunately, any hybrid with a mechanical constituent means that there is no need for the application of complex inverse kinematics to optical or IMU data. These findings are promising for full body capture under 10 milliseconds.

6.2 Mechanical Cost

A full body mechanical system has not been developed for a couple of decades [13], making it difficult to estimate price. To model an entire hand, Sim et al. used angular potentiometers that cost under $5.00 (USD 2023) per unit, along with an Arduino microcontroller that cost approximately $25.00 [60]. This would put estimates of modeling the entire body around $300.00. Note, while optical rotary encoders have greater accuracy than electromagnetic angular potentiometers, they are also more expensive. Therefore, electromagnetic potentiometers should be fully evaluated before examining optical encoders.

6.3 Mechanical Accuracy

Mechanical systems can measure joint angles with less than 1 degree of error using linear or angular potentiometers [22, 61]. To estimate the relative position quality, we use the average male full-arm length of 782 millimeters [62] and the 0.5 degrees of average error achieved by [22] to get approximately 7 millimeters of error. This is far less than the RMSE of 38 millimeters achieved by Atrsaei et al. [56]. In addition to this improved average error, mechanical systems do not suffer from long-term integration error accumulation. Finally, sudden error fluctuations can be filtered out effectively [63], unlike IMUs. Of course, errors from the derivation of instantaneous joint-angular velocity will affect the derivation of joint-angular instantaneous acceleration.

6.4 Resolving Mechanical Suits into Optical Systems

Because mechanical suits only provide relative data, obtaining absolute data requires an optical system. As such, full body occlusions will interrupt absolute measurements. Note, potentiometers can achieve sampling rates comparable to those of IMUs. Therefore, a slower optical refresh rate can limit the mechanical suit, as mechanical samples that occur between the slower optical frames cannot be globally resolved. Despite these flaws, mechanical systems provide stronger relative measurements and thus occlusion resistance.

There are two general methods to resolve discrepancies between the relative position and joint angle data from the mechanical and optical systems. The first approach is to correct mechanical data with superior optical data, improving average relative and absolute accuracy, but reducing the consistency of relative readings during evolving occlusions. The second approach is to treat the entire mechanical suit as a rigid body and resolve it into the optical system unaltered. Although average relative and absolute data will be worse, relative data will be more consistent during evolving occlusions. Given the strengths of relative measurements over absolute measurements and the importance of spatial consistency over accuracy [19, 6], the second approach seems more promising in terms of spatial behavior.

Because the mechanical suit is treated as a rigid body in the second approach, only a total of three noncollinear markers are needed to resolve the entire suit into the optical system. Therefore, there is potential to create a short-circuit mechanism that halts the processing of additional markers once three per user have been located, thus lowering optical latency [35]. Essentially, the second approach always processes mechanical data, but allows for the processing of markers to be minimized. Conversely, the first resolving approach implies it is only necessary to process mechanical data during occlusions, while markers are always processed. Note, under optimal conditions, optical latency exceeds mechanical latency. Therefore, the minimization of optical latency should be prioritized by taking the second resolving approach.

If the second approach is taken, users should be made aware of the sources and behaviors of relative and absolute measurements. Regardless of the resolving approach taken, absolute position will have a bias towards the set of visible markers. Optimally, the set of visible markers will be evenly distributed. But if only a small portion of the body is visible, accuracy for occluded points distal to the visible section will have lower accuracy.

6.5 Ergonomic Alternatives

Unfortunately, mechanical systems suffer from poor ergonomics. Exoskeletons can be cumbersome and restrictive. There are ergonomic alternatives, although with major caveats.

Stretch sensors are fabricated from conductive elastomers that relay deformation data via changes in capacitance [64]. In other words, they capture motion by directly modeling soft tissue deformation. Stretch sensors are significantly more ergonomic than exoskeletons. Unfortunately, their recovery delay was found to be 40 to 200 milliseconds by Nguyen et al. in [65]. This is the time fabric takes to relax to its original shape after a deforming force has ceased. Additionally, data processing is significantly more complex than exoskeletal systems [66].

Biomedical goniometers and torsiometers also provide a more ergonomic mechanical system. Goniometers measure angular rotation in two orthogonal planes, such as radial and ulnar deviations, and wrist extensions and flexions. Torsiometers measure angular twisting on a single axis, such as forearm pronation and supination. Both sensors have similar working mechanisms. A wire with integrated strain sensors relays joint angle information as it bends or twists. Shiratsu et al. found biomedical goniometers to consistently provide error of less than 6 degrees [67]. Although more consistent than IMU angle readings [57], it is much less accurate than exoskeletons [61]. These devices take approximately 8 milliseconds to acquire measurements [68], meaning it could meet our latency threshold. Unfortunately, these biomedical devices cannot model more than 2 planes of motion, and thus they cannot fully model ball joints, such as the shoulder. Finally, each unit costs thousands of dollars, thus modeling all joints will cost more than a commercial optical system.

Fiber optic strain sensors measure the change in light transmissions when an optical fiber is deformed [69]. Both D'Mello et al. and Nishiyama et al. showed that fiber optic sensors can provide under one degree of joint angle accuracy [70, 71]. Unfortunately, the manufacture of fiber optic sensors is much more expensive than electromagnetic potentiometers [72], and thus they should only be explored if there is no other way to improve ergonomics.

6.6 Improving Traditional Mechanical Suits

As we can see, all the alternatives to electromagnetic mechanical systems have major caveats. To improve the ergonomics of a traditional mechanical system, we can place the potentiometers directly on the skin, in a form fitting manner, similar to [63], thus avoiding the restrictions of an exoskeleton. Unfortunately, placing potentiometers on the skin at the axis of the joint makes them more susceptible to soft tissue artifacts. Fortunately, some artifacts are predictable and can be modeled by an external system, specifically, the deformation of muscles that cause the motion of the joint. We could use a stretch sensor to model the soft tissue deformation of each joint. Alternatively, optical markers can be placed on the body portions surrounding the joint, but distal to the muscles causing the motion, thus avoiding the artifacts caused by muscle deformation. The values of the potentiometers would then be mapped to the joint angle values provided by the optical system. The potentiometer could then properly compensate for soft tissue artifacts. This method of mapping known angles is similar to a method used to calibrate goniometers [73]. This approach will be more difficult for complex ball-and-socket joints such as the shoulder which has forward, upward, and twisting dimensions, all of which interact with one another. For example, the potentiometer used for measuring a forward motion will be affected by the deformation of muscles causing an upward or twisting motion, which are primarily measured via other potentiometers.

Note that to secure sensors and markers, a common industry standard is to use straps and morph suits. Unfortunately, [4] found that morph suits are uncomfortable for extended periods of time. Furthermore, this type of approach is vulnerable to mechanical slippage, which is the shifting of the sensor in relation to the skin [74]. The only alternative is to glue or tape the sensors directly to the skin.

Compared to IMUs, mechanical systems can emulate optical position with greater accuracy and less fluctuation, while having lower latency and cost. Furthermore, they have the potential to provide passive haptic feedback. The challenge of foregoing exoskeletons and attaching potentiometers directly, while maintaining accuracy, is of primary importance for future research.

7 ADDITIONAL INSIGHTS

Before summarizing our findings, wireless networks and the creation of control mechanisms must be briefly considered.

7.1 Network and Latency

To enable user freedom of motion, both IMU and mechanical systems must avoid the use of cumbersome wired setups. Although a full discussion of optimal wireless practices is beyond the scope of this article, a few things should be noted. Under optimal conditions, both Bluetooth and Wi-Fi connections can achieve a one- way transmission latency of under 4 milliseconds [75, 76]. Minimal latency and deviation can be facilitated by hardware design. Firstly, a dedicated network should be used. The congestion of external network traffic was shown by [48] to result in 23.4 milliseconds of latency with a standard deviation of 11 milliseconds for a 2.4GHz 802.11g public WiFi network. Secondly, the airspace itself can become congested with wireless traffic, thus causing packet loss and latency. The HTC Vive system, an inertial-optical hybrid, is susceptible to this problem, as each sensor has its own wireless connection. To free airspace and enable higher user counts, each personal hardware system should limit its wireless transmission volume. One method would be to wire all IMUs into a single on-body unit that would then consolidate the readings and transmit them to a central off-body computer via one wireless connection. Wired USB connections contribute only tenths of a millisecond of delay [77].

7.2 Control Mechanism Creation Software

Cooking data was a term used by Skogstad et al. to refer to the process of mapping motion data to the production of sound [4]. Essentially it is the creation of control mechanisms. Skogstad et al. built a custom application to convert motion data into Open Sound Control (OSC) format, which is designed for real-time sound control. Then, for sound rendering, this data was streamed to Max/MSP, which is a visual programming language for the creation of interactive multimedia systems via the manipulation of audio, MIDI, and video files. Skogstad et al. found Max/MSP to be unsuitable for cooking data. Motion data was also streamed to MVN studio, an Xsens software, to provide visual user feedback. Turchet et al. noted the importance of synchronizing audio and visual feedback with minimal latency [1]. Using separate applications for audio and visuals would complicate this, while also potentially overburdening a single PC [4]. Skogstad et al. noted the potential benefit of the development of single software to complete all cooking.

Game engines provide a suitable starting point. Game engines are software frameworks that provide tools to render visuals and audio, simulate physics, and map user inputs to in-game mechanisms in real-time. Crucially, they can import mocap data, with existing support for VR games and gesture recognition. Additionally, game engines such as Unity or Unreal have integrated sound engines such as FMOD or Wwise. Sound engines are middleware used to create immersive and interactive soundscapes for video games. These sound engines have proven their temporal suitability, having been used to develop the games Rocksmith and Guitar Hero respectively, which rely on synchronized audio and visuals. A handful of applications from academia have leveraged game engines to create musical cooking environments, although they aimed to leverage simple controller-based VR hardware, not full body data [1]. Simultaneously we must achieve a high virtuosity ceiling. To these ends, and given the limited history of embodied synthesis, a musician's pre-existing skills should be leveraged, such as their experience with DAWs. A digital audio workstation (DAW) is a desktop software that is used in the fields of music production and sound engineering as a centralized environment to record, edit, mix, and produce digital audio files.

Despite the capabilities of sound engines and DAWs, Palumbo et al. noted, while creating a virtual reality environment for audio synthesis, that simply inserting a virtual desktop application into a virtual space fails to fully leverage the emerging possibilities of extended reality [78]. Furthermore, given the importance of experimentation during music creation in VR [54], a cooking environment should streamline the creation and alteration of control mechanisms. The demands of hardware management itself can stifle creativity [17], underlining the importance of creating an environment that minimizes the turnaround time between the formulating of a new mapping idea and its implementation. The creation of an environment for streamlined yet expressive embodied synthesis cooking is a novel challenge. While further efforts are required, building upon the capabilities of game engines and DAWs can reduce development costs.

8 SUMMARY, DISCUSSION, AND FUTURE RESEARCH

8.1 Summary of User Experience Analysis

We started by characterizing the user's experiences of different mocap systems and the metrics that they directly sample. To minimize latency and error, we should directly sample metrics that users find intuitive. The simulation of the imparting of energy into the instrument serves to emulate acoustic instruments. To these ends, IMUs can directly acquire absolute linear and rotational acceleration. Providing a coordinate system common to all users allows for direct and novel musical collaboration. Optical systems can directly acquire absolute global position. We found that humans can accurately determine the location of their limbs in relation to each other, but they struggle to determine the location of their limbs in an absolute coordinate system, implying that relative motion is more intuitive than absolute. Moreover, we can easily control our joints separately or in an integrated manner, yet we find it challenging to isolate motion along a given vector in 3D space. To capitalize on these findings, mechanical systems can be used to directly acquire joint angles. Of all non-hybrids, the mechanical system provides the most intuitive directly sourced metric.

To enable creative freedom and high virtuosity, all relevant metrics should be provided in a timely manner. Unfortunately, no relevant systems directly sample relative position, absolute or relative linear velocity, relative linear acceleration, relative orientation, absolute or relative rotational velocity, or relative rotational acceleration. Fortunately, the efficiency of a mechanical-optical system allows time for the calculation of these metrics. Determining the user's experience of these indirectly calculated readings is worthy of further investigation. Note that there are devices that can directly sample joint-angular velocity and acceleration [79, 80], although an investigation into their potential is only necessary if the joint-angular velocity and acceleration derived from an angular potentiometer is found by users to be insufficient.

8.2 Summary of Performance and Cost Analysis

Because each system provides a distinct experience for the user, we explored the potential of hybrid systems, which allows for the direct sampling of multiple metrics. Furthermore, we evaluated the performance and cost of each system, while also exploring the ability of constituent systems to compensate for each other's weaknesses.

Optical latency is sufficiently low. One user's position can be processed in around 5 milliseconds [35]. Commercial systems have approximately 0.2 millimeters of positional error and 0.1 millimeters of standard deviation, and roughly 0.2 degrees of angular error and 0.1 degrees of standard deviation [27]. To address high costs, we can employ smartphone cameras. This will require the synchronization of differing frames rates of given phone models, which in turn necessitates the determination of effective frame rates of any given phone, as these may differ from the manufacturer specifications [43]. Fortunately, the RGB sensors of smartphones allow the use of wavelength methods to identify markers, thus avoiding the downsides of temporal sequence IDs.

To fully address occlusions, we must place more hardware directly on the user. IMU systems seem to be stuck in the balancing act between the accuracy and computational complexity of their filtering methods, resulting in both a high latency and positional error. Processing the position of one user takes roughly 20 milliseconds [58]. Using a non-linear Kalman filter on an IMU, Atrsaei et al. achieved an RMSE of 38 millimeters, with regular error fluctuations that approached 100 millimeters [56]. Sfalcin et al. found that the Xsens suit had an RMSE of 24.5 degrees, with a maximum error of 76.9 degrees [57].

Conversely, mechanical systems provide lower error and latency. Angular potentiometers can achieve as little as 0.5 degrees of average error [22], which then results in around 7 millimeters of positional error, using the average length of a male arm. Finally, sudden error fluctuations can easily be filtered out effectively, unlike IMUs [63]. Modeling one user takes approximately 3 milliseconds, excluding the application forward kinematics [60]. Finally, potentiometers can cost under $5.00 (USD 2023) per unit, while IMUs capable of filtering and integration cost hundreds per unit. As such, IMUs should only be used for raw acceleration readings, while mechanical systems should be used for handling occlusions.

8.3 Future Research Recommendations

Given the poor ergonomics of exoskeletal mechanical systems, our primary future research goal will be to forgo the use of exoskeletons, while maintaining low error and latency. Attaching potentiometers directly to the skin requires us to compensate for soft tissue deformation. This could be done via the use of an external reference, such as an optical system. If both challenges of smartphone integration and mechanical ergonomics are handled, we can further research mechanical-optical hybrids. Specifically, we should investigate how the frequency and magnitude of occlusions, and the resultant accuracy, is impacted by the number of cameras and users in a capture space. Additionally, we must determine how frequently full body occlusions occur. This information will allow us to determine if one additional smartphone camera per user is tenable. Further prototyping and user feedback will be required to hone the optimal system, similar to acoustic instruments, which have evolved for thousands of years [1].

> "The progress of Musical XR research is inexorably bound to that of XR hardware and software technologies."
> [1]

REFERENCES

[1] Luca Turchet, Rob Hamilton, and Anil Camci. 2021. Music in Extended Realities. *IEEE Access* PP, (January 2021), 1–1. https://doi.org/10.1109/ACCESS.2021.3052931

[2] David Wessel and Matthew Wright. 2001. Problems and Prospects for Intimate Musical Control of Computers. *Computer Music Journal* 26, (December 2001). https://doi.org/10.1162/014892602320582945

[3] Juan Pablo Wachs, Mathias Kölsch, Helman Stern, and Yael Edan. 2011. Vision-based hand-gesture applications. *Commun. ACM* 54, 2 (February 2011), 60–71. https://doi.org/10.1145/1897816.1897838

[4] Ståle Skogstad, Kristian Nymoen, Yago Quay, and Alexander Jensenius. 2011. OSC Implementation and Evaluation of the Xsens MVN suit. *Proceedings of New Interfaces for Music Expression* (January 2011).

[5] Ståle Andreas van Dorp Skogstad. 2014. Methods and Technologies for Using Body Motion for Real-Time Musical Interaction. Doctoral thesis. Retrieved July 1, 2024 from https://www.duo.uio.no/handle/10852/38311

[6] Alessandro Filippeschi, Norbert Schmitz, Markus Miezal, Gabriele Bleser, Emanuele Ruffaldi, and Didier Stricker. 2017. Survey of Motion Tracking Methods Based on Inertial Sensors: A Focus on Upper Limb Human Motion. *Sensors* 17, 6 (June 2017), 1257. https://doi.org/10.3390/s17061257

[7] Laisi Cai, Ye Ma, Shuping Xiong, and Yanxin Zhang. 2019. Validity and Reliability of Upper Limb Functional Assessment Using the Microsoft Kinect V2 Sensor. *Appl Bionics Biomech* 2019, (2019), 7175240. https://doi.org/10.1155/2019/7175240

[8] Andreas Bergsland and Robert Wechsler. 2016. Turning movement into music: Issues and applications of the MotionComposer, a therapeutic device for persons with different abilities. *SoundEffects* 6, (November 2016), 23–47. https://doi.org/10.7146/se.v6i1.24911

[9] Serdar Kucuk and Z. Bingul. 2006. Robot Kinematics: Forward and Inverse Kinematics. https://doi.org/10.5772/5015

[10] C. Pizzolato, M. Reggiani, L. Modenese, and D. G. Lloyd. 2017. Real-time inverse kinematics and inverse dynamics for lower limb applications using OpenSim. *Comput Methods Biomech Biomed Engin* 20, 4 (March 2017), 436–445. https://doi.org/10.1080/10255842.2016.1240789

[11] P. Nogueira. 2012. Motion Capture Fundamentals A Critical and Comparative Analysis on Real-World Applications. 2012. . Retrieved July 1, 2024 from https://www.semanticscholar.org/paper/Motion-Capture-Fundamentals-A-Critical-and-Analysis-Nogueira/051361b591b4f7d981d5a31f67ee353104c86119

[12] Ståle Skogstad, Kristian Nymoen, and Mats Høvin. 2011. Comparing Inertial and Optical MoCap Technologies for Synthesis Control. *Proceedings of the 8th Sound and Music Computing Conference, SMC 2011* (January 2011).

[13] Gypsy Motion capture system technology explained. Retrieved July 2, 2024 from https://metamotion.com/gypsy/gypsy-motion-capture-system-mocap.htm

[14] Andrea Ancillao, Erwin Aertbeliën, and Joris De Schutter. 2021. Effect of the soft tissue artifact on marker measurements and on the calculation of the helical axis of the knee during a gait cycle: A study on the CAMS-Knee data set. *Human Movement Science* 80, (December 2021), 102866. https://doi.org/10.1016/j.humov.2021.102866

[15] Andy Hunt, Yo Dd, Marcelo Wanderley, and Matthew Paradis. 2002. The Importance of Parameter Mapping in Electronic Instrument Design. *Journal of New Music Research* 32, (May 2002). https://doi.org/10.1076/jnmr.32.4.429.18853

[16] Nick Collins, Chris Kiefer, Zeeshan Patoli, and Martin White. 2010. Musical Exoskeletons : Experiments With A Motion Capture Suit. (June 2010). https://doi.org/10.5281/ZENODO.1177749

[17] Ståle Skogstad, Kristian Nymoen, Yago Quay, and Alexander Jensenius. 2012. Developing the Dance Jockey System for Musical Interaction with the Xsens MVN Suit.

[18] Sara Stančin and Sašo Tomažič. 2022. Recognizing Solo Jazz Dance Moves Using a Single Leg-Attached Inertial Wearable Device. *Sensors (Basel)* 22, 7 (March 2022), 2446. https://doi.org/10.3390/s22072446

[19] Warren G. Darling, Brian M. Wall, Chris R. Coffman, and Charles Capaday. 2018. Pointing to One's Moving Hand: Putative Internal Models Do Not Contribute to Proprioceptive Acuity. *Front Hum Neurosci* 12, (May 2018), 177. https://doi.org/10.3389/fnhum.2018.00177

[20] John M. Barden, Robert Balyk, V. James Raso, Marc Moreau, and Keith Bagnall. 2005. Repetitive pointing to remembered proprioceptive targets improves 3D hand positioning accuracy. *Hum Mov Sci* 24, 2 (April 2005), 184–205. https://doi.org/10.1016/j.humov.2005.03.004

[21] Victor Khabie, Michael C. Schwartz, Andrew S. Rokito, Maureen A. Gallagher, Frances Cuomo, and Joseph D. Zuckerman. 1998. The effect of intraarticular anesthesia and elastic bandage on elbow proprioception. *Journal of Shoulder and Elbow Surgery* 7, 5 (September 1998), 501–504. https://doi.org/10.1016/S1058-2746(98)90202-6

[22] Xiaochi Gu, Yifei Zhang, Weize Sun, Yuanzhe Bian, Dao Zhou, and Per Ola Kristensson. 2016.: An Inexpensive and Lightweight Mechanical Exoskeleton for Motion Capture and Force Feedback in VR. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, May 07, 2016. ACM, San Jose California USA, 1991–1995. https://doi.org/10.1145/2858036.2858487

[23] Laura Mikula, Sofia Sahnoun, Laure Pisella, Gunnar Blohm, and Aarlenne Zein Khan. 2018. Vibrotactile information improves proprioceptive reaching target localization. *PLoS One* 13, 7 (July 2018), e0199627. https://doi.org/10.1371/journal.pone.0199627

[24] Joseph Rovan, Marcelo Wanderley, and Shlomo Dubnov. 1997. Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance. (December 1997).

[25] Joanne M. Wagner, Catherine E. Lang, Shirley A. Sahrmann, Qungli Hu, Amy J. Bastian, Dorothy F. Edwards, and Alexander W. Dromerick. 2006. Relationships between Sensorimotor Impairments and Reaching Deficits in Acute Hemiparesis. *Neurorehabil Neural Repair* 20, 3 (September 2006), 406–416. https://doi.org/10.1177/1545968306286957

[26] Robert J. K. Jacob, Linda E. Sibert, Daniel C. McFarlane, and M. Preston Mullen. 1994. Integrality and separability of input devices. *ACM Trans. Comput.-Hum. Interact.* 1, 1 (March 1994), 3–26. https://doi.org/10.1145/174630.174631

[27] Matt Topley and James G. Richards. 2020. A comparison of currently available optoelectronic motion capture systems. *Journal of Biomechanics* 106, (June 2020), 109820. https://doi.org/10.1016/j.jbiomech.2020.109820

[28] Daniel Pacheco de Queiroz, J. Gomide, and A. Araújo. 2012. Evaluation of Real Time Tracking Methods for an Open Source Motion Capture System. 2012. . Retrieved July 2, 2024 from https://www.semanticscholar.org/paper/Evaluation-of-Real-Time-Tracking-Methods-for-an-Queiroz-Gomide/3be0d51abb2eca39e322de712b4aa73838aa1b86

[29] Alfred Franz, Alexander Seitel, Dominique Cheray, and Lena Maier-Hein. 2018. Polhemus EM tracked Micro Sensor for CT-guided interventions. *Medical Physics* 46, (December 2018). https://doi.org/10.1002/mp.13280

[30] David Tsaregorodtsev, Alexander Chugunov, and N. Petukhov. 2019. *Ultra-Wideband Motion Capture Radio System*. https://doi.org/10.1109/FarEastCon.2019.8934111

[31] Minghui Zhao, Tyler Chang, Aditya Arun, Roshan Ayyalasomayajula, Chi Zhang, and Dinesh Bharadia. 2021. ULoc: Low-Power, Scalable and cm-Accurate UWB-Tag Localization and Tracking for Indoor Applications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3 (September 2021), 140:1-140:31. https://doi.org/10.1145/3478124

[32] Fernando J. Álvarez Franco. 2019. 17 - Fundamentals of Airborne Acoustic Positioning Systems. In *Geographical and Fingerprinting Data to Create Systems for Indoor Positioning and Indoor/Outdoor Navigation*, Jordi Conesa, Antoni Pérez-Navarro, Joaquín Torres-Sospedra and Raul Montoliu (eds.). Academic Press, 335–351. https://doi.org/10.1016/B978-0-12-813189-3.00017-4

[33] Sebastian Thrun. 2008. Simultaneous Localization and Mapping. In *Robotics and Cognitive Approaches to Spatial Mapping*, Margaret E. Jefferies and Wai-Kiang Yeap (eds.). Springer, Berlin, Heidelberg, 13–41. https://doi.org/10.1007/978-3-540-75388-9_3

[34] Riccardo Monica and Jacopo Aleotti. 2022. Evaluation of the Oculus Rift S tracking system in room scale virtual reality. *Virtual Reality* 26, 4 (December 2022), 1335–1345. https://doi.org/10.1007/s10055-022-00637-3

[35] Real-Time Latency of VICON. *Biomch-L*. Retrieved July 2, 2024 from https://biomch-l.isbweb.org/forum/biomch-l-forums/biomch-l-1988-2010/19069-real-time-latency-of-vicon

[36] You Li, Julien Moreau, and Javier Ibanez-Guzman. 2023. Emergent Visual Sensors for Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems* 24, 5 (May 2023), 4716–4737. https://doi.org/10.1109/TITS.2023.3248483

[37] Ashley Chey Vincent, Haley Furman, Rebecca C. Slepian, Kaitlyn R. Ammann, Carson Di Maria, Jung Hung Chien, Ka-Chun Siu, and Marvin J. Slepian. 2022. Smart Phone-Based Motion Capture and Analysis: Importance of Operating Envelope Definition and Application to Clinical Use. *Applied Sciences* 12, 12 (January 2022), 6173. https://doi.org/10.3390/app12126173

[38] Nobuyasu Nakano, Tetsuro Sakura, Kazuhiro Ueda, Leon Omura, Arata Kimura, Yoichi Iino, Senshi Fukashiro, and Shinsuke Yoshioka. 2020. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Frontiers in Sports and Active Living* 2, (2020). Retrieved June 8, 2023 from https://www.frontiersin.org/articles/10.3389/fspor.2020.00050

[39] Yebin Liu, Juergen Gall, Carsten Stoll, Qionghai Dai, Hans-Peter Seidel, and Christian Theobalt. 2013. Markerless Motion Capture of Multiple Characters Using Multiview Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 11 (November 2013), 2720–2735. https://doi.org/10.1109/TPAMI.2013.47

[40] Mobile Vendor Market Share Worldwide. *StatCounter Global Stats*. Retrieved July 2, 2024 from https://gs.statcounter.com/vendor-market-share/mobile/worldwide/

[41] Plask Motion: AI-powered Mocap Animation Tool. Retrieved July 2, 2024 from https://plask.ai/docs/44

[42] Luis Almeida, Paulo Menezes, and Jorge Dias. 2020. Interface Transparency Issues in Teleoperation. *Applied Sciences* 10, 18 (September 2020), 6232. https://doi.org/10.3390/app10186232

[43] Kengo Teramoto, Ryo Mukainakano, Hiroki Watanabe, H. Hashizume, and M. Sugimoto. 2019. 3D tracking using smartphones for a marker-based optical motion capture system. 2019. . Retrieved July 2, 2024 from https://www.semanticscholar.org/paper/3D-tracking-using-smartphones-for-a-marker-based-Teramoto-Mukainakano/7400c87c5779005779d037d32b572aa1252153ae

[44] Yibo Zhu, Xin Lei, Ken Wang, and Zongfu Yu. 2019. Compact CMOS spectral sensor for the visible spectrum. *Photonics Research* 7, (September 2019), 961. https://doi.org/10.1364/PRJ.7.000961

[45] Shoji Tominaga, Shogo Nishi, and Ryo Ohtera. 2021. Measurement and Estimation of Spectral Sensitivity Functions for Mobile Phone Cameras. *Sensors* 21, 15 (January 2021), 4985. https://doi.org/10.3390/s21154985

[46] David Flam, Daniel Queiroz, Thatyene Ramos, Arnaldo Araújo, and Joao Gomide. 2009. OpenMoCap: An Open Source Software for Optical Motion Capture. *2010 Brazilian Symposium on Games and Digital Entertainment* 0, (October 2009), 151–161. https://doi.org/10.1109/SBGAMES.2009.25

[47] Chang Peng, Qianqian Cai, Mengyue Chen, and Xiaoning Jiang. 2022. Recent Advances in Tracking Devices for Biomedical Ultrasound Imaging Applications. *Micromachines (Basel)* 13, 11 (October 2022), 1855. https://doi.org/10.3390/mi13111855

[48] Geise Santos, Johnty Wang, Carolina Brum, Marcelo M. Wanderley, Tiago Tavares, and Anderson Rocha. 2021. Comparative Latency Analysis of Optical and Inertial Motion Capture Systems for Gestural Analysis and Musical Performance. In *International Conference on New Interfaces for Musical Expression*, June 01, 2021. . https://doi.org/10.21428/92fbeb44.51b1c3a1

[49] Yue Li, Dongdong Weng, Dong Li, and Yihan Wang. 2019. A Low-Cost Drift-Free Optical-Inertial Hybrid Motion Capture System for High-Precision Human Pose Detection. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, October 2019. 75–80. https://doi.org/10.1109/ISMAR-Adjunct.2019.00034

[50] Marit A. Zandbergen, Jasper Reenalda, Robbert P. van Middelaar, Romano I. Ferla, Jaap H. Buurke, and Peter H. Veltink. 2022. Drift-Free 3D Orientation and Displacement Estimation for Quasi-Cyclical Movements Using One Inertial Measurement Unit: Application to Running. *Sensors* 22, 3 (January 2022), 956. https://doi.org/10.3390/s22030956

[51] Haoran Yang, Yong Wang, Huiqing Wang, Yunde Shi, Lifeng Zhu, Yujin Kuang, and Yuan Yang. 2022. Multi-Inertial Sensor-Based Arm 3D Motion

Tracking Using Elman Neural Network. *Journal of Sensors* 2022, (May 2022), e3926417. https://doi.org/10.1155/2022/3926417

[52] Hong-Yu Zhao, Zhelong Wang, Qin Gao, Mohammad Hassan, and Abdulhameed Alelaiwi. 2015. Smooth estimation of human foot motion for zero-velocity-update-aided inertial pedestrian navigation system. *Sensor Review* 35, (September 2015). https://doi.org/10.1108/SR-03-2015-0032

[53] Matti Raitoharju and Robert Piché. 2019. On Computational Complexity Reduction Methods for Kalman Filter Extensions. *IEEE Aerospace and Electronic Systems Magazine* 34, (October 2019), 2–19. https://doi.org/10.1109/MAES.2019.2927898

[54] Michele Geronazzo and Stefania Serafin (Eds.). 2023. *Sonic Interactions in Virtual Environments*. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-031-04021-4

[55] Xiaoping Yun and E.R. Bachmann. 2007. Design, Implementation, and Experimental Results of a Quaternion-Based Kalman Filter for Human Body Motion Tracking. *Robotics, IEEE Transactions on* 22, (January 2007), 1216–1227. https://doi.org/10.1109/TRO.2006.886270

[56] Arash Atrsaei, Hassan Salarieh, Aria Alasty, and Mohammad Abediny. 2018. Human Arm Motion Tracking by Inertial/Magnetic Sensors Using Unscented Kalman Filter and Relative Motion Constraint. *Journal of Intelligent & Robotic Systems* 90, (May 2018). https://doi.org/10.1007/s10846-017-0645-z

[57] Bianca Sfalcin, Xiaoxu Ji, Adrian de Gouw, Jim Potvin, and Joel Cort. 2019. Optical and inertial motion capture joint angle comparison using Jack™. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (November 2019), 1137–1141. https://doi.org/10.1177/1071181319631153

[58] MVN Hardware Overview. Retrieved July 5, 2024 from https://base.movella.com/s/article/MVN-Hardware-Overview?language=en_US

[59] MoCapLab documentation data types. Retrieved June 12, 2023 from https://www-hs--augsburg-de.translate.goog/mocap/lernen/dokumentation/datentypen.html?_x_tr_sl=auto&_x_tr_tl=en&_x_tr_hl=en-US&_x_tr_pto=wapp

[60] Donghyun Sim, Yoonchul Baek, Minjeong Cho, Sunghoon Park, A. S. M. Sharifuzzaman Sagar, and Hyung Seok Kim. 2021. Low-Latency Haptic Open Glove for Immersive Virtual Reality Interaction. *Sensors* 21, 11 (January 2021), 3682. https://doi.org/10.3390/s21113682

[61] Yeongyu Park, Jeongsoo Lee, and Joonbum Bae. 2014. Development of a finger motion measurement system using linear potentiometers. In *2014 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, July 2014. IEEE, Besacon, 125–130. https://doi.org/10.1109/AIM.2014.6878066

[62] Gurminder Singh, Ashish Singla, and Gurvinder Virk. 2016. Modeling and Simulation of a Passive Lower-Body Mechanism for Rehabilitation.

[63] Christin Büttner, Thomas L. Milani, and Freddy Sichting. 2021. Integrating a Potentiometer into a Knee Brace Shows High Potential for Continuous Knee Motion Monitoring. *Sensors* 21, 6 (January 2021), 2150. https://doi.org/10.3390/s21062150

[64] Youngsu Cha, Jeonggyu Seo, Jun-Sik Kim, and Jung-Min Park. 2017. Human–computer interface glove using flexible piezoelectric sensors. *Smart Materials and Structures* 26, (May 2017), 057002. https://doi.org/10.1088/1361-665X/aa6b64

[65] Thao Nguyen and Michelle Khine. 2020. Advances in Materials for Soft Stretchable Conductors and Their Behavior under Mechanical Deformation. *Polymers* 12, 7 (July 2020), 1454. https://doi.org/10.3390/polym12071454

[66] Steven Lao, Hamza Edher, Utkarsh Saini, Jeffrey Sixt, and Armaghan Salehian. 2019. A Novel Capacitance-Based In-Situ Pressure Sensor for Wearable Compression Garments. *Micromachines* 10, 11 (November 2019), 743. https://doi.org/10.3390/mi10110743

[67] Angela Shiratsu and H. J. C. G Coury. 2003. Reliability and accuracy of different sensors of a flexible electrogoniometer. *Clinical Biomechanics* 18, 7 (August 2003), 682–684. https://doi.org/10.1016/S0268-0033(03)00110-4

[68] DataLOG for Portable Data Acquisition and Monitoring. *Biometrics Ltd*. Retrieved July 5, 2024 from https://www.biometricsltd.com/datalog.htm

[69] Eric Fujiwara, Yu Tzu Wu, Luiz E. Silva, Hugo E. Freitas, Stenio Aristilde, and Cristiano M. B. Cordeiro. 2019. Optical Fiber 3D Shape Sensor for Motion Capture. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, March 2019. IEEE, Osaka, Japan, 933–934. https://doi.org/10.1109/VR.2019.8798003

[70] Yannick D'Mello, James Skoric, Lea Moukarzel, Siddiqui Hakim, and David V. Plant. 2019. Wearable Fiber Optic Sensors for Biomechanical Sensing via Joint Angle Detection. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2019. IEEE, Berlin, Germany, 32221–3225. https://doi.org/10.1109/EMBC.2019.8857061

[71] Michiko Nishiyama and Kazuhiro Watanabe. 2009. Wearable Sensing Glove With Embedded Hetero-Core Fiber-Optic Nerves for Unconstrained Hand Motion Capture. *IEEE Transactions on Instrumentation and Measurement* 58, 12 (December 2009), 3995–4000. https://doi.org/10.1109/TIM.2009.2021640

[72] Xuhui Zhang, Chunyang Wang, Tong Zheng, Haibin Wu, Qing Wu, and Yunzheng Wang. 2023. Wearable Optical Fiber Sensors in Medical Monitoring Applications: A Review. *Sensors (Basel, Switzerland)* 23, 15 (August 2023). https://doi.org/10.3390/s23156671

[73] PASPORT Goniometer Probe • PS-2138. *PASCO scientific*. Retrieved July 5, 2024 from https://www.pasco.com/products/sensors/pasport/ps-2138

[74] Daniel Xu. 2015. Wearable Motion Capture Stretch Sensors. 2015. . Retrieved August 6, 2023 from https://www.semanticscholar.org/paper/Wearable-Motion-Capture-Stretch-Sensors-Xu/6c08d8182e5a7ed5ca88bf9b82380f65d6a5aed3

[75] Vineet Gokhale, Mohamad Eid, Kees Kroep, R. Venkatesha Prasad, and Vijay Rao. 2021. Towards Enabling High-Five Over WiFi. https://doi.org/10.48550/arXiv.2111.01634

[76] Carles Gomez, Joaquim Oller, and Josep Paradells. 2012. Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. *Sensors* 12, 9 (September 2012), 11734–11753. https://doi.org/10.3390/s120911734

[77] Raphael Wimmer, Andreas Schmid, and Florian Bockes. 2019. On the Latency of USB-Connected Input Devices. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (*CHI '19*), May 02, 2019. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300650

[78] Michael Palumbo, Alexander Zonta, and Graham Wakefield. 2020. Modular reality: Analogues of patching in immersive space. *Journal of New Music*

Research (January 2020). Retrieved July 5, 2024 from https://www.tandfonline.com/doi/abs/10.1080/09298215.2019.1706583

[79] T. L. Briggs, J. D. Campbell, and O. J. Tassicker. 1971. An angular-velocity transducer with applications in the measurement and control of rotation. *J. Phys. E: Sci. Instrum.* 4, 3 (March 1971), 240. https://doi.org/10.1088/0022-3735/4/3/019

[80] Hao Zhao and Hao Feng. 2015. A Novel Permanent Magnetic Angular Acceleration Sensor. *Sensors* 15, 7 (July 2015), 16136–16152. https://doi.org/10.3390/s150716136