# AudioMove: Applying the Spatial Audio to Multi-Directional Limb Exercise Guidance

CHENGSHUO XIA*, Xidian University, China
TIAN MIN*, Keio University, Japan
YUTA SUGIURA, Keio University, Japan

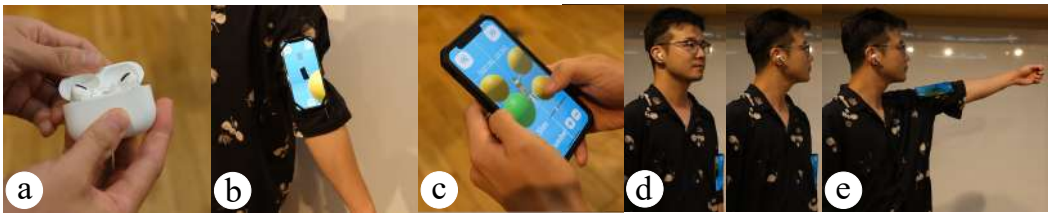


Fig. 1. By wearing earphones (a) and attaching a smartphone to the limb (b), the users engage in customized (c) multi-directional limb exercises guided by spatial audio. The inertial measurement unit (IMU) within the earphones combines users' head motions with spatial audio perception, thereby enhancing their capacity to locate the target (d). Subsequently, users can use their limbs to reach the target, realizing non-visual motion guidance.

Guiding users with limb exercise can assist in muscle training or physical recovery. However, traditional vision-based methods often require multiple camera angles to help users understand the motions and require them to be within the range of the screen. Therefore, we propose a non-visual system that can guide users with multiple-directional limb motions utilizing spatial audio, *AudioMove*, with commercial-off-the-shelf (COTS) devices (i.e., smartphones and earphones). The proposed system addresses the challenge of conveying directional information encompassing multiple planes in real-time. We conduct a mixed-method user study to evaluate the effectiveness of the system with three methods combining motion data with spatial audio perception. Additionally, a user interface is built to collect users' comments. The results conclude that spatial audio guidance could create a natural, pervasive, and non-visual exercise training solution in daily life.

CCS Concepts: • **Human-centered computing** → **Ubiquitous computing**.

Additional Key Words and Phrases: Auditory Feedback, Interface, Guidance, Exercise System, Spatial Audio

**ACM Reference Format:**
Chengshuo Xia, Tian Min, and Yuta Sugiura. 2024. AudioMove: Applying the Spatial Audio to Multi-Directional Limb Exercise Guidance. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 244 (September 2024), 26 pages. https://doi.org/10.1145/3676489

*Authors contributed equally and listed alphabetically. Chengshuo Xia is the corresponding author.

Authors' Contact Information: Chengshuo Xia, xiachengshuo@xidian.edu.cn, Xidian University, Guangzhou, China; Tian Min, welkinmin@keio.jp, Keio University, Yokohama, Japan; Yuta Sugiura, sugiura@keio.jp, Keio University, Yokohama, Japan.


ACM 2573-0142/2024/9-ART244
https://doi.org/10.1145/3676489

## 1 Introduction

Motion exercise has played an essential role in various fields, including restoring physical function [5], keeping fitness [17], or strength training [39]. To better assist the users in exercise training, in recent years, exercise systems have been iterated to accommodate better end-user needs, including remote systems, incorporating exergames, immersive virtual reality, and so on [87]. Such systems are moving toward greater ubiquity and specialization to bring users better results and training experiences.

One of the essential functions of an exercise system is motion guidance. The most classic training process is that the user watches the motion performed, imitates the motions, the system captures the movement trajectory through external devices, and provides feedback to the user for adjustment and improvement. Such a paradigm has been intensively applied both in commercial and research work. For example, the mobile applications [66], the researches of MotionMA [78] Physio@Home [74], SleeveAR [71], GuideBand [77], and PoseAsQuery [24]. However, vision-based guidance systems undoubtedly offer more intuitive learning information and potentially better performance in accuracy. Relying on visual cues requires more significant space to accommodate screens and motion capture devices (e.g., Microsoft Kinect [13]). Users also need to gaze at the screen and remain within the camera's range at all times during the learning process. With the development of head-mounted displays, virtual reality (VR) [16], and augmented reality (AR) [71] have to some extent alleviated the screen requirements. However, head-mount displays are still constrained by relatively lower accessibility and lab-based setups to enable precise motion capture.

Researchers have been actively seeking various ways and alternative modalities to assist users in understanding and performing correct motions. To get rid of visual information reliance, systems based on other modalities have been extensively investigated, for example, the haptic [20] and auditory [29, 46, 64]. While both haptic and auditory modalities can to some extent substitute for vision, they differ in terms of information transmission efficiency and the required equipment for implementation (i.e., speakers, headphones, or wearable actuators). Non-visual guidance makes it accessible in a wider range of scenarios, such as outdoors, places with limited space, the user lying down, or the user simply unwilling to use the vision. However, as a medium for information transmission, haptic and auditory is less intuitive compared to vision, lacking the clarity of an instant understanding of the designated motion. Moreover, they offer fewer adjustable attributes (i.e., volume, pitch, or the intensity of vibration), making it challenging to make users understand complex motions in a short period [48]. As a result, current non-visual feedback is usually used as an enhancement combined with other modalities [3, 83].

Taking the limb movement in exercises as an example, the difficulty in designing such a non-visual system is that, the limb's movement is usually in three dimensions. Intuitively conveying the positional information with a single modality of audio is challenging, most notably in simultaneously determining the direction and height of movements. However, the spatiality of audio offers a potential solution to non-visual exercise systems. Spatial audio can bring sufficient cues on direction as a feedforward to users [47], and has been utilized for navigation [23, 41]. However, it has not been tested to guide the motion of the body so far. Compared to navigating users to walk towards a direction, guiding specific limb movements through spatial audio may require users to possess a higher ability of spatial audio perception and comprehension of the relative position between their limbs and the audio source. In this paper, to realize the guidance for the multi-directional limb exercise, we employed the characteristics of spatial audio and proposed a system that guides the limb exercise through interactive auditory feedback. The system is based on accessible mobile devices (a smartphone and earphones) without any other external infrastructure, aiming for a more seamless and portable solution as the growing research interest on systems implemented with daily COTS devices [4, 54, 83]. The main contributions of the paper are as follows:

- A system employing spatial audio is proposed to fill the blank of non-visual multi-directional limb exercise guidance.
- The effectiveness and performance of the proposed system with different approaches of combining motion data with spatial audio perception were evaluated. Users' experiences on using the spatial audio for motion exercises were collected through semi-structured interviews.

## 2 Related Work

### 2.1 Virtual-and-Real Exercise Tool

Various systems have been proposed with different techniques to assist the exercise training better. Exergames have been popular in fusing entertainment and exercise both in commercial and research works [10, 42, 59, 69, 75]. For example, exercising with fitness equipment [60], setting a virtual teacher [25], or augmentation with instructive information [36]. In addition, utilizing the sense of body ownership, the user could naturally follow the movement of the virtual avatar to conduct the motions, such as Just Follow Me [84], Onebody [28], Tai-Chi training [14] among others. Li et al. [45] also modeled the awareness of such inconsistency of body ownership intervention to further support the VR-based exercise system. Without VR/AR equipment, the systems improving video tutorials also arouse interest. For example, ReactiveVideo [15] integrated the interaction between the playback of exercises and the user's movement speed. YouMove [1] allowed the exercise authoring process in training and the PoseAsQuery enables a start frame variation [24]. Since such systems provide an interface between the user's movement and video, external devices to capture the users' postures are normally required (such as vision sensors [39, 44]). Other than letting users move by themselves, several works investigated the methods of assisting the users via external equipment as well, like using the electrical muscle stimulation [55, 56, 58], mechanic equipment [77] or robotics [21, 22].

Previous works mainly rely on visual information to facilitate users' motion. A few systems focused on employing vibration or auditory to offer additional user experience in training [29, 64]. Vibration uses the haptic sensations of users to lead their movements [49]; for example, the TIKL [46] utilized eight vibrotactile actuators to improve human motion learning. Generally, such a vibration-based system requires several actuators to be worn on, limiting its ubiquity [33]. As for auditory, it has wider and more ubiquitous cases because of the accessibility and persuasiveness of speakers and earphones. However, the audio is usually used as an assistive tool to support conveying visual information, such as the verbal instruction in VoLearn [83], voice reminder in [72] and [73]. The auditory-based exercise system requires further exploration, especially in getting free from other modalities to meet the broader set of scenarios.

### 2.2 Guiding the User with Auditory Information in Exercise Training

To successfully guide the user in an exercise motion, the feedback from the user's movement is essential [67]. The feedback process could be categorized into real-time feedback and post-feedback. The real-time feedback contains an instruction process (i.e., feedforward). It offers users instant information, allowing them to understand where their movements should go. For example, the multiple perspectives in Physio@Home [74], instructive sign [11], and arrow information in LightGuide [70]. Employing the metaphor is also an effective approach that is more intuitive and easy to understand, like mapping the user's movement error into the variation of another object's shape or attribute [35, 36]. In terms of auditory-based real-time feedback, some work employed the pitch of the audio as an effective metaphor to support users to reach the correct position [48, 83].

Post-feedback in auditory-based systems typically provides the user with an evaluation of their performance, e.g., the score in [71] and the emoji in [76]. The verbal information could be a straightforward and effective way to convey the assessment. COPD-Trainer [72] introduced a mobile application for chronic pulmonary obstructive disease patients' rehabilitation. Similarly, VoLearn [83] utilized verbal information to offer the improvement strategy regarding motion speed. Casamassima et al. [12] used the audio bio-feedback message to Parkinson's patients to analyze their gait patterns in rehabilitation.

### 2.3 Spatial Audio-based Design

Spatial audio is a great tool to simulate the perception of auditory spatiality [19]. Users could recognize the direction from the auditory and correspond it to a physical location. Thus, spatial audio has been successfully applied to many entertainment systems to increase the sense of immersion and reality of users [50], such as in video or VR games [9, 65]. Such a feature has also been utilized in navigation systems. For example, for visual-impairment people, spatial audio is feasible to guide the user in wayfinding [86], or it can be combined with a white cane to navigate the user [68]. To support the utilization of spatial audio, virtual spatial audio renders such as the Ambisonics [88] have been developed. Hu et al. [31] used the Ambisonics combined with a camera and laser sensor to build an environmental perception system for blind amputees. They also actively designed a wearable locating system with a web camera [30]. With only using earphones for spatial audio, many products have emerged such as AirPods [2], Microsoft Soundscape [51], and Bose [7]. These products follow the concept of acoustic augmented reality (AAR), which coordinates the user's motion with spatial audio. Ear-AR [85] introduced an indoor AAR design that integrates head motion with spatial audio for indoor localization.

Spatial audio has been widely used due to its directional characteristics, while its exploration in exercise systems is still limited. It is only used in a few rehabilitation systems, and it is often used as a rendering tool to stimulate the user to recover from hearing diseases. For example, to recover from hearing impairment [81], postural control dysfunction [80], and autism spectrum disorder [37].

### 2.4 Spatial Audio-based Motion Guidance System

| | | Information Modality | |
|---|---|---|---|
| | | **Visual** | **Non-visual** |
| **Guiding Limb's Motion** | **Single Direction** | Single Perspective; Projection-based [1, 15, 18, 25, 29, 72, 77] | Audio; Vibrotactile [35, 48, 51, 73, 74] |
| | **Multi-direction** | Single Perspective; Projection-based [26, 75, 79, 85] | ***Our Work*** |

Fig. 2. Summary of exercise systems with different information modality and exercise types.

We determine the main characteristics of such an auditory-based exercise system within its design space. Figure 2 shows the summary of the existing work's taxonomy in terms of information modality employed and the types of exercise guided.

*Information Modality.* Currently, the majority of exercise systems are dependent on visual input because of its simplicity and intuitiveness. As the general scheme of motion guiding process proposed in [78], the users are required to receive demonstrated motion from the professional-end and intimate the motion before the feedback is transmitted to improve their performance until they reach the goal. The information modality indicates the method of providing the user with demonstrated motion and the way to send feedback. Many systems use visual information to address both processes [1, 14, 15, 24, 25, 28, 71, 74, 76, 78, 84]. To meet the non-visual needs, other modalities such as auditory and vibration are also employed [29, 33, 46, 49, 64, 72, 73]

*Limb Exercise Type.* In this paper, we focused on one of the specified types of limb exercise. This type of motion exercise is usually used in strength training of limb muscles, rehabilitation, or physical therapy [6, 18, 43]. With this motion as the target, the existing exercise system can be categorized into supporting single and multi-directional limb movement. The former only involves the movement of a limb in a single axis. For example, in the dumbbell weight training for the upper arm, users raise their forearm along the axis in a transverse plane [40]. While the multi-direction limb exercises realize the movement of limbs and joints in multiple planes simultaneously. In weight training, it helps the user to achieve multi-directional training of muscles [63]. For systems supporting the multi-directional limb exercise, only a visual-based system has been designed to provide feedback and assistance [74].

Thus, we propose a unique method to fill the blank by utilizing the auditory-based (non-visual) approach and support the multi-directional limb exercise. Specifically, spatial audio was employed to give the directional information, and the feedback would help the users adjust their movement to reach the ultimate target position.

## 3 Design of Using Spatial Audio to Support the Limb Exercise

### 3.1 Background of Spatial Audio

The sense of audio spatiality comes from the temporal and volume difference in how the sound waves interact with human ears: interaural time difference (ITD) and interaural level differences (ILD), helping humans determine the azimuth angle (i.e., the direction) of the audio source. ITD is the temporal delay between sound wave arrival at our left and right ears, helping determine the azimuth angle of low-frequency sounds. However, the threshold for detectable ITD increased faster than exponentially for higher-frequency sounds, making it almost impossible for humans to identify the direction of sound with the temporal differences [27]. Instead, due to the weaker penetration of high-frequency sound, the human head's acoustic shadow will cause a difference in loudness and frequency between the two ears, known as ILD. Theoretically, there exists audio with specific patterns and frequency that can achieve optimal localization [53].

However, the elevation angle of an audio source (i.e., the height) is difficult to localize due to the symmetric alignment of ears [82]. In this case, the difference in frequency caused by sound waves entering the external ear, specifically the pinna, from different angles enables humans to determine the sound's elevation to a limited extent [52]. This phenomenon is known as spectral effects, which can be simulated by applying head-related transfer functions (HRTFs).

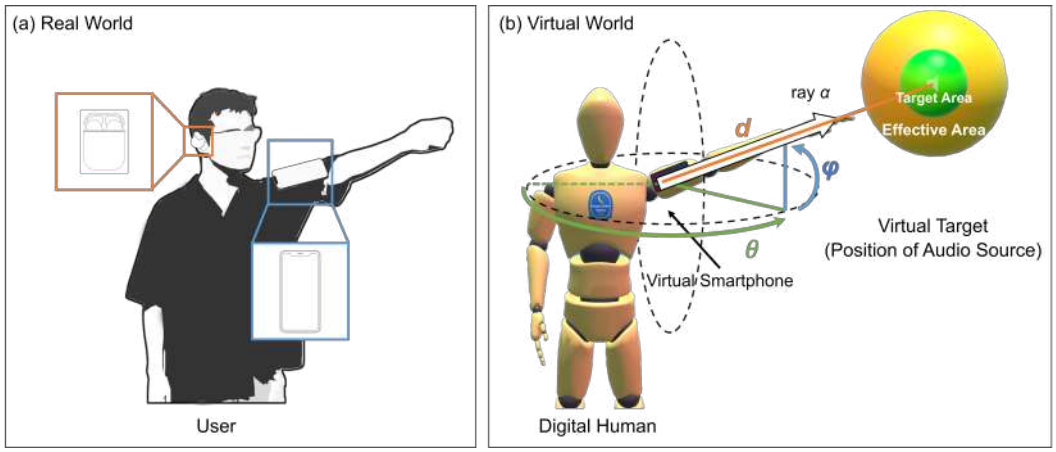


Fig. 3. When users in the real world (a) put on earphones and attach smartphones to their limbs, their auditory perception and limb motions are linked to the system's virtual world (b). The details of the virtual target's position parameters are listed in Table 1.

### 3.2 Synchronizing the real world to the Virtual World

Spatial audio could provide the basis of directional information to users via the spatial perception ability of each individual. The position of an audio source can be represented by a spherical coordinate system. Similarly, we could also characterize a limb motion with the same spherical coordinate system. In this way, we can guide individuals through multi-directional exercises by detecting body limb rotation data in conjunction with spatial audio. We developed a virtual environment with Unity3D to render spatial audio[1] and replicate the rotation of the joint onto a digital human. As shown in Figure 3 (a), a smartphone and earphones embedded with IMUs are used to synchronize users' head and joint rotation to the digital human.

Table 1. Parameters used to describe the position of a virtual target and joint rotation.

| Parameter | Symbol | Description |
|---|---|---|
| Azimuth | $\theta$ | Offset angle from the right in degrees |
| Elevation | $\varphi$ | Offset angle from the horizontal plane in degrees |
| Distance | $d$ | Distance from the center of a virtual target to the origin |
| Pointing Direction (ray) | $\alpha$ | Vector $(\theta, \varphi)$ indicates the pointing direction of body limb |

Virtual targets are set up as the source of spatial audio, as shown in Figure 3 (b). Each virtual target has two layers: a smaller inner sphere as the final designated position of the motion (i.e., *target area*), and a bigger outer sphere (i.e., *effective area*) surrounding the *target area* as guidance, which will be described in detail in the following Section 3.3. In the virtual environment, the *virtual smartphone*, which synchronizes the gyroscope data of the real smartphone attached to the user's limb, is set as the origin of the spherical coordinate. Therefore, the position of a virtual target as well as the pointing direction of the user's limb can be represented with the parameters listed in Table 1.

[1]https://resonance-audio.github.io/resonance-audio/

### 3.3 Process of Reaching the Target Position

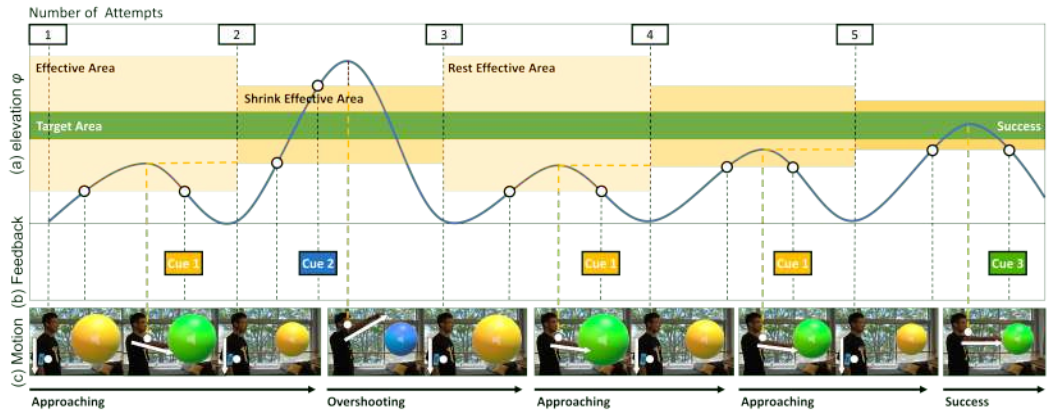


Fig. 4. An example of user approach the *target area* based on the feedback design. As marked on the left side of the figure: (a) Elevation angle ($\varphi$) curve of user's arm. The green and yellow ranges represent *target area* and *effective area* respectively. (b) Auditory feedback cues are given when the user exits the *effective area*. (c) User with a smartphone attached to the right upper arm and the virtual target. The white arrows indicate the direction where the user's arm is pointing to.

Due to individual differences in spatial audio perception, accurately guiding users to reach a target position with their limbs is challenging, especially in determining the height [52]. We designed interactive auditory feedback to enhance the guidance by adaptively and dynamically changing the size of the *effective area* of the virtual target. Figure 4 shows the process of how the user approaches the designated position with the feedback design.

In this example, the user takes five attempts to reach the final position with the right upper arm, and the number of attempts is marked on the top of the figure. The row at the bottom (Figure 4 (c)) shows the user's motions and the virtual targets. The top row in the figure (Figure 4 (a)) shows the curve of elevation angle ($\varphi$) of the user's arm. The green and yellow areas represent the *target area* and *effective area* respectively. The size of the *target area* is fixed, while the size of the *effective area changes* dynamically according to the user's previous attempts. The row in the middle (Figure 4 (b)) shows the timing when the feedback comes up. Whenever a user exists the *effective area*, the system will provide 3 different cues depending on their attempts[2]. The details of the cues are presented in Table 2, where $\varphi_{peak}$ is used to denote the highest point of the elevation trajectory.

Table 2. Auditory cues used in the feedback design.

| Cue | Trigger Condition | Interpretation | System Action |
|---|---|---|---|
| Cue 1 | $\varphi_{peak} \in$ *effective area* | General direction of was correct | Downsize the *effective area* |
| Cue 2 | $\varphi_{peak} >$ *effective area* | The attempt overshot. | Reset the *effective area* |
| Cue 3 | $\varphi_{peak} \in$ *target area* | The attempt was successful | Move to the next position. |

[2]Demonstration available in the corresponding video figure at the ACM Digital Library.

### 3.4 Design Specifications

*3.4.1 Adjust the size of the effective area adaptively.* The size of *target area* and *effective area* of a virtual target can reflect the difficulty of reaching it. A larger *target area* implies that the curve traced by a user's limb is more likely to fall within the successful range. Meanwhile, a larger *effective area* means that users can easily trigger the entering sound effect, indicating they're heading towards the correct direction, which allows them to receive positive feedback from the initial attempt. To realize this, we set the size of the initial state of the *effective area* to be fixed and as large as possible, with its radius approaching the distance between its center and the virtual smartphone ($d$) without affecting the collision volume, which means the sphere is tangent to the origin (virtual smartphone). It is sufficient to maintain a ratio between the distance from the virtual target sphere center to the origin and the radius of the *effective area* to keep the same level of difficulty.

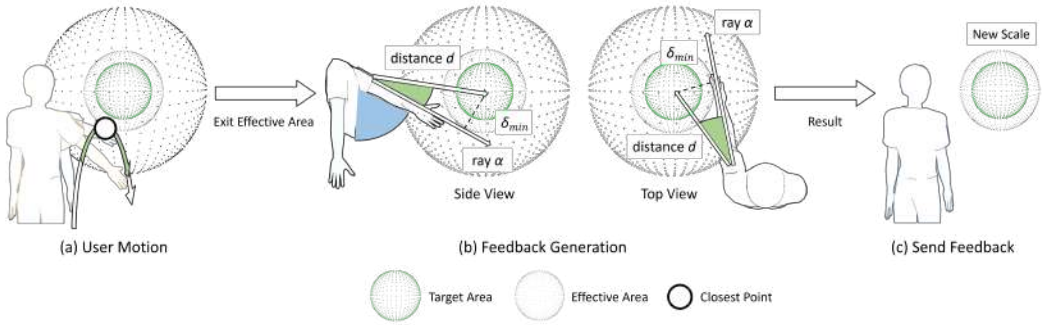


Fig. 5. The principle of size adjustment of the *effective area*. The user conducts a limb exercise of raise up and put down in (a). The white arrow in (a) represents the user limb's trace, while the green arrow represents the trace within the *effective area*. The white dot in (a) marks the closest point ($\delta_{min}$) to the *target area*. (b) Shows the side view and top view of the snapshot when the angle between limb direction and target direction is minimal. After exiting the *effective area*, we calculate the shrank scale of *effective area*, $s$. (c) After that, the new scale takes effect with the feedback cues sent to the user.

In order to realize adaptive and dynamic feedback, we adjust the *effective area* size of the virtual target according to user motion. As shown in Figure 5 (a), when the trajectory drawn by the user's arm intersects with the virtual target sphere, we record the angle ($\delta$), between the ray ($\alpha$) and the direction vector of the virtual target during this process. As the user's arm trajectory exits the *effective area*, the system calculates the radius of the new *effective area* based on the minimum angle recorded, by $radius = \sin(\delta_{min}) \cdot d$, where $d$ is the distance between the virtual smartphone and virtual target's center, equals to the initial radius of the *effective area* (Figure 5 (b) and (c)).

*3.4.2 Calibration.* To address the possibility of user moving or rotating their body while using the system, we introduced a calibration function. Taking the upper arm motion involving shoulder joint rotation as an example, users start the system with their arms naturally hanging down. At this point, the system uses the initial posture of the smartphone as a reference to generate the virtual target in the direction the user is facing. If users move or turn around, the calibration function allows them to adjust the virtual target's relative position based on their current orientation.

### 3.5 Methods Combining Motion Data with Spatial Audio

The perception ability on spatial audio can be further enhanced through dynamic ITD and ILD variations [79]. As more and more mobile devices we carry nowadays are embedded with IMUs, and using COTS devices has been becoming a rising research interest [4, 54, 83], we are able

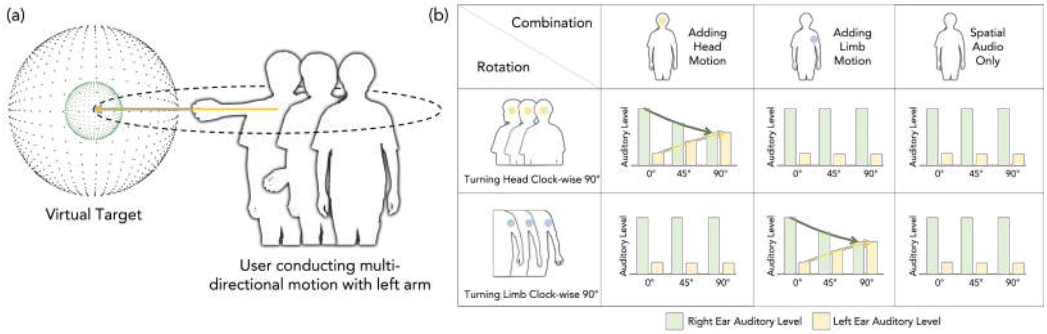


Fig. 6. (a) A virtual target is positioned at the right side of a user, and it's playing spatial audio. (b) The table enumerates methods for combining motion data from different body parts, specifically the head, limb, and none. The rows of the table represent the user performing clockwise rotations of the head or limb by 90 degrees. The bar chart in each cell illustrates the audio levels received by the left and right ears of the user during the process of rotating the head or limb.

to collect the rotational data from multiple sensed points on user's body and apply them to the listener's posture in the virtual world. In the proposed system, there are two mobile devices: a smartphone attached to the user's exercising limb and earphones worn on the head, which result in two approaches that combine motion data to spatial audio perception: *combining head motion* and *combining limb motion*. Additionally, considering neither as applicable serves as a baseline (*spatial audio only*), resulting in three methods.

Figure 6 (a) shows an example: a virtual target is positioned to the user's immediate right, playing spatial audio. As shown in Figure 6 (b), applying the user's head motion data to the listener in the virtual space allows a perceptible balance shift of the audio heard from both ears as the user gradually turns the head to the right by 90 degrees; similarly, applying the user's limb motion data to the listener allows the user to perceive a gradual balancing of the audio as the limb rotates clockwise to the right. No such effects occur under other conditions.

## 4 Spatial Audio-based Limb Exercises Interface

To investigate how users employ the spatial audio-based multi-directional limb exercise system in customized scenarios and collect their comments on this guidance method, we developed a user interface (UI) for evaluation purposes.

**(a) Audio Customization:** Since the main guidance approach is based on audio with spatial characteristics, we allow users to select their preferred music as the audio source shown in Figure 7 (a).

**(b) Training Body Limbs Selection:** Users can click on different limbs of the digital human, and the system will display the names of corresponding body limbs as shown in Figure 7 (b). The virtual smartphone would also be presented after the limb is selected. These locations define where the user will place their smartphone on their body. In the UI, we have defined a total of eight selectable limbs, which are *right upper arm, right forearm, left upper arm, left forearm, right thigh, right shin, left thigh and left shin.*

**(c) Virtual Targets Positions with Different Repetitions and Difficulty:** After the basic settings of music and body limb, we provide an interface to customize the exercise target positions as shown in Figure 7 (c), (d), (e). This opens up personalized training routines. Users could drag the

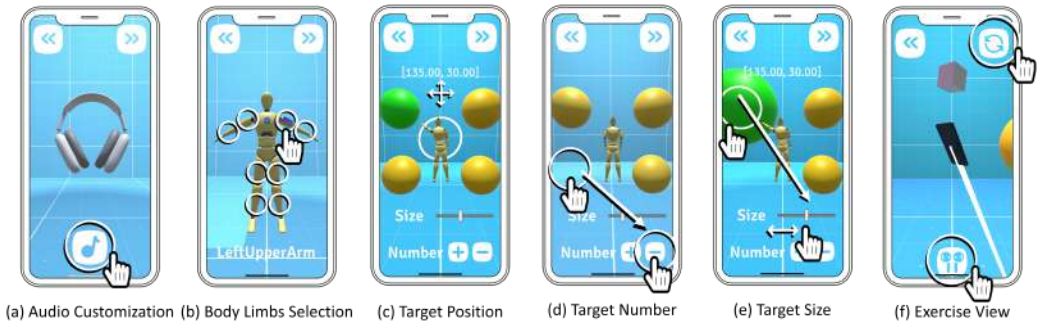


Fig. 7. (a) In the audio customization view, the user can click on the button to select preferred audio. (b) In the body limb selection view, the user can click on a digital human to choose the training limb. (c), (d), (e) Motion target position configuration view. The virtual targets can be clicked to be selected, and the user can perform a series of customized settings, including adjusting their positions, number, and size. (f) shows the exercise view, where the user can click on a button for calibration, and an earphone button to switch among three methods of combining motions to spatial audio.

virtual targets along a spherical coordinate, while the selected limb would move with the dragging target to present the posture of the human body in this target position. An ergonomic constraint of $\theta \in (30°, 150°), \varphi \in (-45°, 45°)$ is applied to the spherical coordinate, considering the normal joint rotation range to prevent users from creating unreachable targets [61, 62]. Moreover, users could decide the number of sphere targets by selecting and clicking the plus or minus button. Users can also configure the size of each target to adjust the difficulty of reaching them. Larger virtual targets (*target area*) imply users can reach them more easily. Conversely, smaller targets need more precise and rigorous limb motion control from the users. After configuring the virtual target positions and before the training, the system will provide users with a preview by playing an animation of the digital human. A verbal prompt will remind users to wear their devices to the selected limb at the same time.

## 5 Evaluation

In order to prove the effectiveness of the proposed system, we suggest three research questions:

- *Question 1* (RQ1): Can spatial audio support conducting multi-directional limb exercises in different motion patterns?
- *Question 2* (RQ2): In the methods of combining motion data and spatial audio suggested in Section 3.5 (head, limb, and spatial audio only), which one of them could provide better experience and accuracy?
- *Question 3* (RQ3): How will the users use and comment on the spatial audio guidance with their customized targets?

We divide the evaluation into two parts. In *Study 1*, we focus on testing the performance and accuracy of the system. In this study, we test the three types of methods combining motion data with spatial audio perception on different motion patterns, in order to answer RQ1 and RQ2. In *Study 2*, we conducted semi-structured interviews with the participants regarding their comments and experience with the UI and the customization process to answer RQ3.

### 5.1 Study 1: Guiding Methods

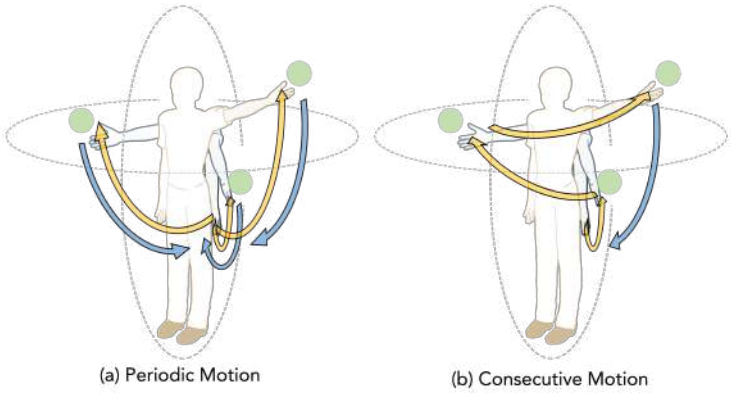


Fig. 8. A user is conducting multi-directional limb motion with the left upper arm in (a) periodic and (b) consecutive ways. The green circles represent the designated positions of the motions. The yellow arrows indicate the direction moving towards the designated positions, while the blue arrow indicates the movement towards the naturally hanging initial state of the user's arm.

*5.1.1 Motion Tested.* A multi-directional limb motion requires repetitive adduction and abduction, flexion, and extension in multiple planes, which is commonly seen in relaxation, weight-training, and rehabilitation [6, 18, 43]. In this study, we take the left upper arm as an example, to examine the feasibility of spatial audio guidance with two types of motion patterns.

As shown in Figure 8, a user is conducting multi-directional upper arm motion targeting three designated positions in periodic and consecutive ways. In the periodic motions, the user moves the arm with a raise-up and put-down pattern, reaching the positions one at a time. In the consecutive motions, the user's limb moves to the target and continues reaching the following position in sequence.

*5.1.2 Participants.* We recruited 24 participants (7 females and 17 males) with an average age of 24.8 (SD = 2.4) from the university. The participants were all right-handed, without hearing problems, and did not have daily exercise routines. Each participant was compensated with 15 dollars.

*5.1.3 Experiment Setup.* In order to answer the research questions RQ1 and RQ2, we validate the guidance with spatial audio for two types of motion patterns: *periodic* and *consecutive*, with three types of method combining motion data with spatial audio perception: *head motion data*, *limb motion data*, and using *spatial audio only* as the baseline. We report the quantitative participants' performance by the number of attempts and errors in degrees, and the qualitative results of participants' experience by semi-structured interviews. We set a between-subject experiment with the *combining motion data method* as an independent variable. Thus, three groups in total were formed with each of 8 people. The groups are labelled with ***Group_Spatial*** (*spatial audio only*), ***Group_Limb*** (*combining limb motion*), and ***Group_Head*** (combining head motion).

For each group, three tasks were designed for evaluating different multi-directional motion patterns under different combining motion data methods. During the tasks, as our participants were all right hand as the dominant hand, to eliminate the effect from different habits of different user dominant hands. A smartphone (iPhone 12 Pro) was attached to their *left upper arm*, and a pair of AirPods Pro were used as the earphones. The IMU data from smartphone and earphones[3]

[3]https://developer.apple.com/documentation/coremotion/cmheadphonemotionmanager

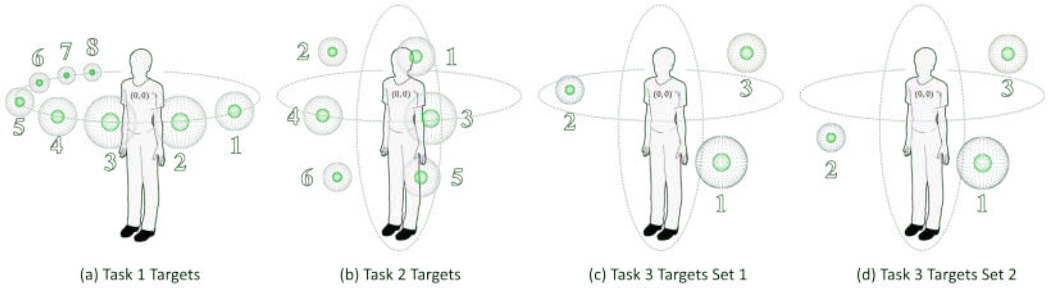


Fig. 9. Three tasks illustration and position of each target. The size of the target looks different as the perspective is different. They are the same size during the evaluation. The inner green sphere refers to *target area*, while the outer sphere refers to *effective area*.

are connected and processed through the application installed. A music clip was used as the audio source [4]. The tasks are described as follows:

**Task 1:** In this task, we investigated the discrepancy between the participants' perceived direction and the actual direction of the audio source. We measured only the azimuth angle error without taking the number of attempts or the task completion time into account. Participants were instructed to point their arms toward the direction they perceived as the audio source. As shown in Figure 9 (a), we sequentially generate eight random positions with the azimuth angle in the range of $\theta \in (30°, 150°)$ on the horizontal plane. Upon the participant's confirmation of their decision, we record the azimuth angle error between the user's indicated direction and the actual direction.

**Task 2:** To explore the impact of three combining motion data methods on user performance under the context of *periodic motion*, as shown in Figure 9 (b), we set six different positions in front of the user, and a virtual target playing spatial audio will appear at these positions in a random order. The virtual target only moves to the next position when the user successfully reaches the *target area* of the previous virtual target. The task concludes when the user identifies all six positions. Specifically, these positions, are denoted in spherical coordinates $(\theta, \varphi)$ as $(60°, 45°)$, $(60°, 0°)$, $(60°, -45°)$, $(120°, 45°)$, $(120°, 0°)$, $(120°, -45°)$.

**Task 3:** Finally, for the *consecutive motions*, two position sets are designed, with each set containing three positions. The virtual target will appear at specified positions in a fixed order, and similarly, it will move to the next position only after the user reaches the previous one. In the context of *consecutive motion*, once the user arrives at a position, there is no need to return to the initial position. Instead, guided by spatial audio and feedback from the virtual target, the user proceeds to the next position until all designated positions. Specifically, two sets denoted as $(120°, -45°)$, $(60°, 15°)$, $(135°, 45°)$ and $(120°, -45°)$, $(60°, -15°)$, $(135°, 45°)$, respectively as illustrated in Figure 9 (c) and (d).

All the virtual target positions set during the evaluation are limited by the system's ergonomic constraint as mentioned in Section 4. User's attempts where the angle error between the user's arm pointing direction and the virtual target direction less than 20 degrees (i.e., the *target area* range), were considered as a success, while the *effective area* was set to maximum, ± 42 degrees as its initial state.

During three tasks, the following data are recorded as the measurement of participants' performance and experience:

[4]Music Clip: https://www.youtube.com/watch?v=Opp9nqiN5m0.

(1) ***Azimuth Angle Error*** between participants' perceived direction and actual virtual target direction in **Task 1**. The average error of eight randomly generated positions was calculated to measure the accuracy of participants' perception accuracy.

(2) ***Number of Attempts*** before reaching the designated position. In **Task 2** and **Task 3**, the number of attempts indicates how much effort was made by participants during the task. A higher number of attempts indicates the participants were challenged in reaching the correct position.

(3) ***Average Angle Error*** for all positions between the user arm's direction and virtual target position's direction when the virtual target is reached in **Task 2** and **Task 3**. It represents how close the participant's limb is to the intended direction. For example, the participant can perfectly conduct the motion if the angle error equals 0.

(4) ***Task Completion Time***. We recorded the task completion time for **Task 2** and **Task 3** using the timer implemented within the software. For the same set of targets, a lower number of attempts should correspond to a reduced task completion time. However, considering the purpose outlined in the RQs, we did not urge participants to complete the tasks as quickly as possible. Instead, we allowed them ample time to perceive and judge the position of the audio source based on the feedback.

(5) ***Workload***. Each participant completed the NASA-TLX [26] on their experience throughout the study. It validates if the proposed method of combining head/limb motion data will increase the physical workload and cognitive load compared to using spatial audio-only.

*5.1.4 Procedure.* After the demographic investigation was obtained from the participants, we introduced the content of the experiment to the participants. We ensured that each participant had a brief period of around 10 minutes to try the system before each task, providing them with familiarity with spatial audio and the feedback design. Subsequently, we conducted the experiments in the order of **Task 1**, **Task 2**, and **Task 3**. After all three tasks were completed, we provided the NASA-TLX questionnaire and conducted a semi-structured interview with the participants to collect the qualitative results. The user study was generally completed within an hour.

### 5.2 Study 1: Results

*5.2.1 Quantitative.* The recorded data was analyzed by the Kruskal-Wallis test with the *combining motion data method* as a variable from *Group_Spatial* to *Group_Head*, considering comparing more than two unmatched groups with normality not assumed [57]. We presented the results of the significant effect of three methods as follows.

For **Task 1**, Figure 10 (a) presents the result of the error of azimuth angle between the participants' perception direction and spatial audio direction from three groups. The Kruskal-Wallis test shows a significant effect of *method* ($x^2$ = 17.540, p < 0.001) among three groups. Post-hoc tests were conducted using Mann-Whitney test with Bonferroni correction showed the significant differences between *Group_Spatial* and *Group_Limb* ($x^2$ = 62.0, p < 0.001), between *Group_Spatial* and *Group_Head* ($x^2$ = 64.0, p < 0.001) and between *Group_Limb* and *Group_Head* ($x^2$ = 56.0, p < 0.05). Using spatial audio with head motion data could perceive the direction better.

For **Task 2**, Figure 10 (b) shows the result of an average number of attempts before participants reach the target positions. A significant effect of group on the number of attempts was revealed ($x^2$ = 8.343, p < 0.05). Post-hoc tests showed a significant effect between the *Group_Spatial* and 3 ($x^2$ = 53.0, p < 0.05) and between the *Group_Limb* and 3 ($x^2$ = 57.5, p < 0.001). In addition, Figure 10 (e) showed the results of the average task completion time in seconds participants spent before reaching each target, which also demonstrated a significant effect of the *method* ($x^2$ = 6.86, p < 0.05) among three groups. Post-hoc tests showed the significant differences between *Group_Spatial* and *Group_Head* ($x^2$ = 51.0, p < 0.05), between *Group_Limb* and *Group_Head* ($x^2$ = 51.0, p < 0.05). These

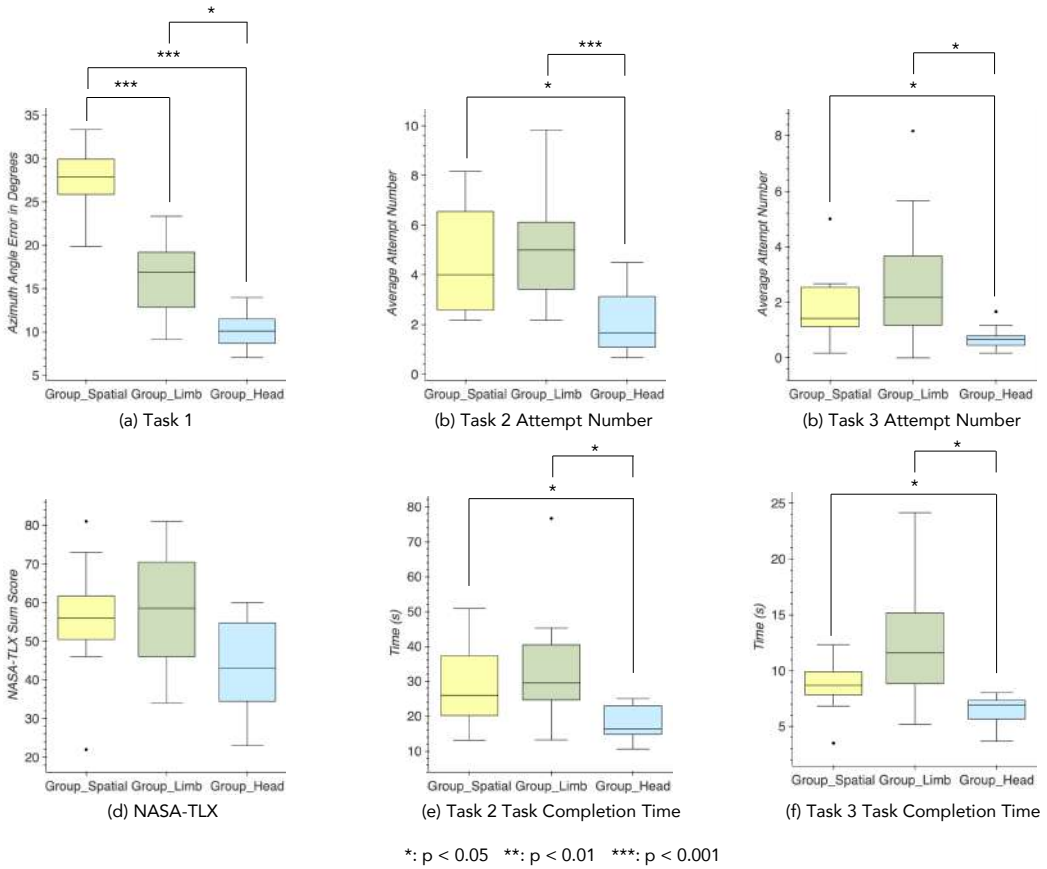


Fig. 10. Result of three tasks and cognitive load investigation.

| | Group_Spatial | | | Group_Limb | | | Group_Head | | |
|---|---|---|---|---|---|---|---|---|---|
| | Avg Attempt | TCT | Avg Error | Avg Attempt | TCT | Avg Error | Avg Attempt | TCT | Avg Error |
| Task 1 | - | - | 26.28 | - | - | 19.91 | - | - | 11.40 |
| Task 2 | 4.56 | 29.35 | 10.31 | 5.27 | 35.12 | 10.74 | 2.12 | 17.88 | 11.12 |
| Task 3 | 1.91 | 8.52 | 15.34 | 2.93 | 12.56 | 16.21 | 0.72 | 6.39 | 12.43 |

Fig. 11. The result of the number of attempts, average task completion time spent to reach each target (seconds), and average error (degrees) in **Task 1**, **Task 2** and **Task 3**.

findings indicate that using spatial audio with head motion data could help the user to minimize the seeking process in periodic motions (**Task 2**). In addition, we reported the average error angle between the participant's limb direction and spatial audio direction when reaching the target (as in Figure 11).

Figure 10 (c) showed the result of **Task 3** on the average number of attempts. It revealed a significant effect of 3 groups ($x^2 = 6.325$, $p < 0.05$). With the same post-hoc tests, significant effects were revealed between the *Group_Spatial* and *Group_Head* ($x^2 = 51.0$, $p < 0.05$) and between the *Group_Limb* and *Group_Head* ($x^2 = 53.0$, $p < 0.05$), which proving the combining head motion

data is still the optimal choice under consecutive motion context. For the task completion time, a significant effect of 3 groups ($x^2 = 7.634$, $p < 0.05$) was found, where the post-hoc test revealed the significant effect between *Group_Spatial* and *Group_Head* ($x^2 = 51.0$, $p < 0.05$) and between *Group_Limb* and *Group_Head* ($x^2 = 53.0$, $p < 0.05$), as shown in Figure 10 (f).

We also conducted the significance test on three groups' NASA-TLX results, and there is no significant effect regarding the cognitive load ($p > 0.05$). In other words, even combining the limb's motion data with the spatial audio could create a new auditory feeling for the user, while not increasing the mental burden.

*5.2.2 Qualitative.* We conducted the semi-structured interview to explore how users think of the spatial audio-based limb exercise experience. Our goal was to reveal to what extent the participants can adapt spatial audio as non-visual guidance, as well as how the feedback design and the combining motion data method contributed to their experience.

The interview scheme was conducted in the form of open-ended questions. Comments from the participants were encoded with thematic analysis and affinity diagramming [8]. Results echoed the main themes of our concerns: the feasibility of spatial audio-based exercise guidance and the comparison among three combining motion data methods.

***Theme 1: Spatial audio-based guidance is considered appealing and convenient by the participants.*** When being asked about the use of spatial audio to guide motions, all three groups of participants' comments can be described using keywords like novelty, fun, or convenience. Many of their comments were related to the process of locating the audio source, such as:

> *"I like to be led by spatial audio to conduct motion because it is relaxing to me that I can just follow the variations of audio to make exercise"* (P24). *"I really enjoy completing the motion tasks in one position to move on to the next position. It gives me a sense of satisfaction"* (P19).

***Theme 2: Combining motion data to spatial audio eases the process of locating, where Group_Head is considered the optimal choice.*** For combining the limb motion data, participants in *Group_Limb* liked this feature when they were seeking direction by rotating the arm. For example:

> *"The idea of aligning your arm to the audio is interesting"* (P8). *"It is much better to watch on a monitor to understand the information"* (P13). *"I feel tracking a balance in my ears."* (P15). *"It is more useful when I want to find a direction"* (P16).

However, *Group_Limb* participants left some comments regarding the negative aspect. The most mentioned point is taking additional steps before conducting the motion. They normally needed to first rotate the limb (as in **Task 1**) to judge the direction and then conduct the motion. Some participants provided the following comments:

> *"I just want to directly move my arm instead of judging the correctness of direction at first."* (P11). *"When doing Task 2 and Task 3, rotating my arm in mid-air makes a weird (auditory) feeling."* (P14)

Despite the method of combining limb's motion data being recognized as fun or interesting, *Group_Limb* do not reveal a significant effect in terms of attempt number of reaching the target compared to *Group_Spatial* in **Task 2**. From their feedback, combining the limb's motion data is advantageous for them in determining the target's direction while the arm is vertically down. On the contrary, it tends to cause confusion when the arm is lifting to reach the target or during arm rotations in mid-air. Compared to *Group_Limb*, we received more positive feedback from participants in *Group_Head*, which is in line with the significant effect between them in all three tasks. They generally perceived this as a more immersive and natural approach:

*"It's a straightforward way to pinpoint it."* (P18). *"I located the target by comparing the volume when turning head quite precisely."* (P20). *"I can quickly reduce the range of search by simply looking around."* (P22).

In summary, we can answer RQ1 and RQ2 with the quantitative and qualitative results. For RQ1, using spatial audio as a guidance can effectively help participants reach the designated positions in different motion patterns, and is considered an appealing and convenient approach by the participants. Taking *Group_Head* as the example, participants take less than 2 attempts before reaching the correct position on average, with an error lower than 15 degrees. For RQ2, combining head motion data with spatial audio perception (*Group_Head*) stands out as the optimal approach. It not only significantly reduces the number of attempts, but also receives more preference and higher comments from the participants.

### 5.3 Study 2: User's Experience on the Spatial Audio Guidance with the Customized Targets

*5.3.1 Participant.* We invited 8 participants (2 females and 6 males) with an average age of 23.2 (SD=1.7) from the same university in *Study 1*. They were all right-handed and without hearing problems. Each participant was compensated with 15 dollars.

*5.3.2 Procedure.* Participants were invited to customize a set of targets using the UI. Before users embarked on the design of their customized training, we provided them with a detailed introduction to the functionality of UI. They were allowed to select their preferred music as the audio source and choose the body limb for training. Participants had control over the position, size, and quantity of sphere targets. Before the exercise, they could preview the motions through an animated representation from a digital human. Once satisfied, participants attached the smartphone and initiated the training with *combining head motion data* method (*Group_Head*) until ending the session on their will.

### 5.4 Study 2: Results

Participants' responses were coded with thematic analysis with the same process used in *Study 1*.

***Theme 1: The function of selecting exercise limb meets participants' needs and they ask for more information displayed.*** When choosing the body limb for training, most participants opted for the same smartphone placement as the upper arm in *Study 1*. P26 and P27 chose different positions, with P26 selecting the right lower arm and P27 setting the left lower leg. All participants believed that choosing the body limb was important for their workout, such as:

*"Training limb selection will meet the need from various of the user."* (P28). *"It would be even more supportive to display muscle name or estimated calories burned."* (P27).

***Theme 2: Participants were inclined to utilize the customized target position feature.*** Most participants showed an interest in customizing target positions and willingly spent time designing their own exercises. They tended to arrange targets on the interface in orderly patterns or create specific symbols. For example, P27 attempted to make his arm trace a star trajectory during the training, while P20 endeavored to create the silhouette of a character with targets of varying sizes. Like:

*"The precision of dragging targets on screen is enough for me."* (P25). *"I just want to try out how my star trajectory would feel like in training."* (P27). *"It would be more helpful if there's an easy way to set the sequence of targets when I'm customizing my training routine."* (P25, P28, P30). *"The constraint adjustment could be opened in some 'advanced' mode like some user might want to bend their limbs to their backs."* (P29).

Interestingly, we observed that during customizing target positions, many participants (P25, P27, P28, P29, P30, P31) intended to place the targets on relatively 'extreme' positions (like the boundary position of the ergonomic constraint), which caused some challenges when they attempted their personalized training routines at the beginning, although they all adapted quickly and were able to reach their targets rapidly within a short period. Like:

> *"While it seemed very intuitive on the target-setting interface, the actual experience was quite different. I couldn't get a concrete sense of where the targets would sound like or how challenging I wanted them to be during the setup."* (P28).

In addition, we also observed that none of the participants adjusted the virtual target size (*target area*) to its maximum, which is the easiest state to find, though they were informed of the relationship between size and difficulty. They tended to visually adjust the target to a suitable size to create patterns they liked. One participant commented as follows:

> *"The test play (during the introduction of the system) built up my confidence. I thought making them smaller wouldn't be a big deal."* (P30).

Overall, for answering RQ3, we could state incorporating music and limb selection into the system garnered positive feedback from all participants, indicating its widespread appeal and potential as a motivating factor for exercise with spatial audio. For the target position configurations, the introduced UI system also demonstrated a capacity to effectively address users' exercise needs while attracting user interest.

## 6 Discussion

Spatial audio, when used for directional guidance, is characterized by its ability to swiftly convey positional information to users. Leveraging the inherent directional perception of the human ears, spatial audio offers users a more intuitive and natural means of spatial orientation, devoid of visual cues. The use of spatial audio merges the guidance information into the music or any other kind of audio, which extends the integration of computing assistance into our everyday lives, making the navigation process more intuitive and engaging. Through our experiments, we explored the feasibility of employing spatial audio to guide limb movements multi-directionally and found that it makes it easier to locate audio positions by including head motions. We aim to justify the position of spatial audio among various motion guidance methods by considering factors including information conveyed, applicable motions, and implementation devices.

### 6.1 Conveying Guidance Information

The efficiency with which a motion guidance system conveys information largely determines the effectiveness of users' learning. This information conveys process involves three processes: *the system provides feedforward* to help users understand movements, *the system captures users' motions through sensors* to measure their performance, and *the system provides feedback* to users to correct their movements. In the three processes, the visual modality holds advantages. Users can observe videos to learn movements while the system captures user motions through a camera and provides visual feedback. Its key characteristic lies in providing users with instant information on how the target motion should be executed at the first glance of the example. In contrast, traditional audio and haptic guidance can only describe how close users are to the target motion using variations in pitch, volume, or the intensity of actuators' vibrations. Users must undergo a gradual searching process with the cues. Spatial audio, to some extent, enables the audio with properties similar to video to give users instant information on the approximate direction. While visual modality undoubtedly dominates in terms of information convey efficiency, from a research perspective,

dissecting and reducing the number of modalities used in a guidance system can be helpful for deeper understanding. The motion guidance system explored in this study is based on an audio application.

### 6.2 Limb Motion and Full Body Motion Guidance

A full-body motion can be decomposed into the movements of individual joints. However, in practice, we must consider the issue of information conveyed, as mentioned above, that different information modalities have limits on the amount of information they can effectively convey. Visual methods can quickly enable users to mimic the motions they see, while auditory information, whether in terms of pitch, volume variation, or verbal instructions, requires users to spend more time understanding. This study utilizes spatial audio to efficiently convey the position of targets in space, aiming to increase the information-carrying capacity of the auditory method. We explored the use of the method at various levels, including single joint rotation, periodic, and continuous motions, to enhance the system's capability over limb motion guidance (Section 5.1). The question of whether the spatial audio can be extended to full-body motion guidance is worth exploring. We believe that adding more auditory cues on top of spatial audio can support multiple joint rotations, but this would also result in more complex rules and higher learning costs.

### 6.3 Motion Guidance with COTS Devices

Within a system, the configuration of the three processes of information convey determines the *types of motions that the system can accommodate* and the *devices required for implementation*. For example, providing feedforward through video requires a display, capturing user posture through IMUs requires wearable devices, and providing feedback through haptic requires actuators. When designing systems, researchers aim to integrate the devices serving as information carriers as much as possible for simplicity. With the increasing miniaturization and affordability of electronic components and sensors, users have more choices of COTS devices. The smartphones we carry every day are integrated with more types of sensors and more powerful computational resources. Using workout smartphone holders, users can secure the phone firmly to their limbs, allowing the smartphone to utilize its IMU to detect the motion of the attached limbs (Figure 1). The proposed system outlines a basic prototype of spatial audio-based limb motion guidance, where smartphones and earphones with built-in IMUs are sufficient. However, when designing a more sophisticated system, smartwatches can be used as IMUs to reduce the potential impact of the smartphone's weight. The weight of a smartphone could be ignored in scenarios such as strength training but should be a critical consideration when applying it to rehabilitation [18, 87].

## 7 Design Implications

In this section, we conclude the findings throughout the system design and the research questions proposed in the evaluation, providing insights into spatial audio-based motion guidance and interaction systems design. We summarize the design implications in Figure 12 and elaborate on them in the following.

### 7.1 Implication 1: Combining Limb Motion: intriguing and interesting scheme but may cause confusion.

From the result, participants in *Group_Limb* expressed intrigue with the approach of combining limb motion data with spatial audio perception in the **Theme 2** of *Study 1*. They were able to experience the sensation of 'listening to the audio with their arms' in the virtual environment. This approach extends and enhances the way of users' auditory perceptions, transcending the limitations of the natural body [32]. It has great potential for new training methods. For instance, in

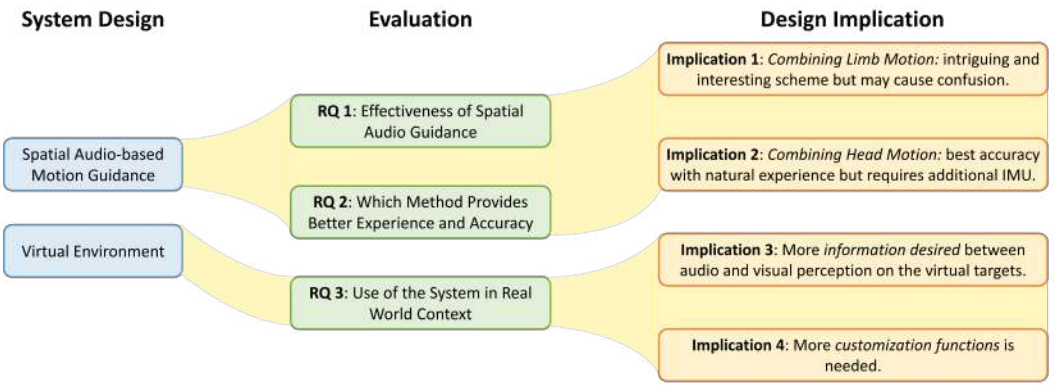


Fig. 12. Implications drawn on the design of a spatial audio-based motion guidance system.

exercises or rehabilitation motions involving joint circumduction [34], we can transfer the rotation of the neck searching for a sound source into the circumduction of limb joints.

However, *Group_Limb* did not show a significant decrease in the number of attempts during **Task 2** and **Task 3**. The primary challenges stem from two key factors. Firstly, a discrepancy exists between users' perceived direction and the actual direction of their limbs moving. This misalignment hinders users from accurately directing their arm movements according to their perception. Secondly, users occasionally faced confusion from the audio variations when moving their limbs in the air. Unintended rotations during limb motions contributed to perplexing audio variations. Despite prior familiarization with the system, participants struggled to swiftly adapt to this pattern or synchronize their auditory sensations within a limited period.

### 7.2 Implication 2: Combining Head Motion: best accuracy with natural experience but requires additional IMU.

On the contrary, *Group_Head* presents a more seamless and natural experience. From *Study 1*, users demonstrated quicker coordination between head and limb motions with this approach, requiring fewer attempts to locate the virtual target. This method has the additional requirement of an IMU-embedded device (i.e., AirPods Pro). However, as we anticipate a growing trend of embedding IMUs in more and more daily wearable electronics, we continue to advance. This trend opens up the prospect of widespread applications for this method. While head rotation serves as a feasible auxiliary solution for locating targets, it has been proved that dynamic ITD and ILD contribute to spatial audio perception [79]. It helps users locate the target's position and further helps them complete the limb exercise more efficiently. We did not intend this functionality to compel users to rotate their heads during motion to achieve precise positioning constantly. At the outset of the design, we aimed to avoid imposing an additional burden on users. Figure 10 alleviates the concerns that *Group_head* did not demonstrate a significant increase in workload while improving the performance.

### 7.3 Implication 3: More information is desired between audio and visual perception of the virtual targets.

We incorporated functions enabling users to design each virtual target's position to accommodate individual preferences. In *Study 2*, while the results generally affirm the usability of the UI, we uncover that users encountered challenges in forming expectations on the position and size of the

target they configured. Users struggle to correlate the on-screen position of the virtual target with the corresponding absolute world motion. While they can visually perceive the size of the virtual target on the UI, they lack a concrete sense of how large the virtual target is when trying to reach it. Despite our intention to provide an intuitive 3D representation of the virtual world, users less familiar with spatial audio still struggled to establish target position and size expectations. This underscores a perceptual gap between visual and auditory modalities. This finding could provide valuable insights for future scenario designs, emphasizing the need for features such as auditory previews to bridge the connections between visual and audio or the utilization of real-world motion data as the configuration input.

Furthermore, due to its non-visual design, while the system is capable of providing guidance, it faces challenges in offering real-time monitoring during exercise. This limitation hinders effectively forming self-awareness or embodiment for users as demonstrated in other systems [38, 71, 76]. Consequently, it may be necessary to introduce a more detailed preview during the configuration phase before the exercise session starts. Alternatively, reintroducing audio variations like pitch during the exercise could be considered based on user needs.

### 7.4 Implication 4: More customization functions is needed.

From *Study 2*, we noticed the customization of virtual targets is appealing to the participants. In addition to the virtual target's position, size, and quantity, spatial audio offers various other aspects that can be tailored to enrich users' exercise routines. In the current design, the virtual target serving as an audio source remains stationary at a fixed position, providing location information to users until they reach it successfully. However, dynamic movements of virtual targets, such as rotating around the user, transitioning from left to right, or varying speed, represent potential interactive features. We can investigate users' perceptual abilities regarding the trajectories of virtual targets and leverage this insight to create novel motion guidance techniques and new auditory sensations.

## 8 Limitation & Future Work

The contribution of this paper lies in the application of spatial audio, proposing a non-visual solution to support multi-directional motion guidance. We validate the effectiveness and feasibility of this approach through the user study. However, a few aspects remain unexplored.

***Comparing With Visual Method***: We acknowledge the intuitiveness of visual-based solutions and do not seek to demonstrate that spatial audio can outperform visual-based solutions in terms of accuracy or serve as a complete substitute. Hence, as in the *Study 1*, we opt to narrow the scope of this study to concentrate on accomplishing spatial audio guidance through diverse methods. Similarly, for feedback design, auditory cues, pitch variations, and verbal cues have proven effective in numerous systems [3, 30, 30], which are also not the focal points of evaluation. Still, comparing the performance of spatial audio guidance to visual-based methods is intriguing. Research focusing on the cognitive load and the level of user acceptance comparison can be conducted in the future.

***Motion Selection and Expansion*:** In our evaluation, the left upper arm was chosen as the focus. We chose a relatively simple motion as the first step to validate the effectiveness of the proposed system, which is common in related work of motion guidance [1, 58, 71, 83]. Additionally, according to the proposed method, all major limbs of the body can be engaged in exercises, including the feedback designed as shown in Figure 13. However, for exercise systems that can be potentially employed in various limbs [38, 83], transferring conclusions to other limbs requires careful consideration of human factors. For the proposed system, the consideration lies in the varying ranges of motion for different joints, a factor that can be appropriately adjusted within the customization options of the UI.


Fig. 13. The system could support non-visual motions driven by various joints in scenarios including (a) forearm exercises outdoors, (b) leg exercises while reading, and (c) leg exercises while lying down.

***Real-world Context Utilization***: While we believe that COTS-based systems are beneficial for users' daily use and can promote their exercise routines, as evidenced by the positive feedback from *Study 2*, we did not, however, require participants to holding weights, perform aerobics exercises or dances combining a set of target positions. Thus lacking certain investigation on integrating spatial audio-based motion guidance into real-world activities. In addition, only the immediate results in this paper were explored. Exploring the long-term learning effects generated by using spatial audio guidance and the in-the-wild usage of users is also worth studying.

***Participant Number in the Study***: We employed a between-subjects approach to mitigate the potential bias stemming from participants becoming increasingly familiar with the spatial audio perception during the study, thereby ensuring the independence of the results from any learning effects. This becomes more pronounced with the three distinct scenarios we had for integrating spatial audio with motion, each encompassing three tasks. The number of participants involved in each group might not be extensive. However, the number is still in line with the basic requirement in similar projects on motion guidance to demonstrate a result [1, 15]. Nevertheless, we acknowledge that an increasing number of participants could further consolidate our findings or potentially reveal more insights.

***Ground Truth of the Limb Motion Angles***: In the study, we assessed whether users directed their limbs toward specified directions within a given range rather than precisely quantifying the angle. For the studies' purposes, ensuring the consistency of the target positions across all participants was paramount, which could be achieved through the calibration process outlined in Section 3.4.2 before each session. However, whether the orientation of the virtual smartphone accurately reflects the orientation of users' limbs in the physical world remains unverified with a ground truth, which could be validated in future work through motion capture techniques.

## 9 Conclusion

We presented using spatial audio to guide the users in conducting limb exercises, particularly filling the blank of supporting multi-directional motions without visual approach. The method maintains the advantages of a non-visual system based on COTS devices, which are portable and can be adapted to a wider range of scenarios. The system's feasibility was tested through a mixed-method user study. We evaluated three methods combining motion data to the spatial audio perception, among which combining head motion data results in the most effective way in terms of less number of attempts taken and better accuracy. Based on our findings, we provide design implications and potential ways to realize a more natural and ubiquitous spatial audio guidance with broader application scenarios.

## Acknowledgments

The work was partially supported by Research Funds for the Central Universities Grant (XJSJ23109), JST PRESTO (Grant Number JPMJPR2134), and KGRI Challenge Grant.

## References

[1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: enhancing movement training with an augmented reality mirror. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. ACM, New York, NY, USA, 311–320.

[2] Apple. 2023. Airpods. https://www.apple.com/airpods/.

[3] Riku Arakawa and Hiromu Yakura. 2021. Mindless Attractor: A False-Positive Resistant Intervention for Drawing Attention Using Auditory Perturbation. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15.

[4] Riku Arakawa, Hiromu Yakura, Vimal Mollyn, Suzanne Nie, Emma Russell, Dustin P. DeMeo, Haarika A. Reddy, Alexander K. Maytin, Bryan T. Carroll, Jill Fain Lehman, and Mayank Goel. 2023. PrISM-Tracker: A Framework for Multimodal Procedure Tracking Using Wearable Sensors and State Transition Information with User-Driven Handling of Errors and Uncertainty. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 156 (jan 2023), 27 pages. https://doi.org/10.1145/3569504

[5] Mobolaji Ayoade and Lynne Baillie. 2014. A novel knee rehabilitation system for the home. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, New York, NY, USA, 2521–2530.

[6] David G Behm, Anthony J Blazevich, Anthony D Kay, and Malachy McHugh. 2016. Acute effects of muscle stretching on physical performance, range of motion, and injury incidence in healthy active individuals: a systematic review. *Applied physiology, nutrition, and metabolism* 41, 1 (2016), 1–11.

[7] Bose. 2023. Bose. https://www.bose.com/en_us/landing_pages/developer-portal.html.

[8] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806 arXiv:https://doi.org/10.1080/2159676X.2019.1628806

[9] James Broderick, Jim Duggan, and Sam Redfern. 2018. The importance of spatial audio in modern games and virtual environments. In *2018 IEEE Games, Entertainment, Media Conference (GEM)*. IEEE, 1–9.

[10] Qinglei Bu, Xiaoyi Cheng, Fan Yang, Jie Sun, Limin Yu, and Ying Hou. 2022. A Computer Game-based Tangible Upper Limb Rehabilitation Device. In *Proceedings of the 10th International Conference on Human-Agent Interaction*. 309–313.

[11] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13.

[12] Filippo Casamassima, Alberto Ferrari, Bojan Milosevic, Laura Rocchi, and Elisabetta Farella. 2013. Wearable audio-feedback system for gait rehabilitation in subjects with Parkinson's disease. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*. ACM, New York, NY, USA, 275–278.

[13] Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang. 2011. A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in developmental disabilities* 32, 6 (2011), 2566–2570.

[14] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for physical tasks in virtual environments: Tai Chi. In *IEEE Virtual Reality, 2003. Proceedings.* IEEE, Long Beach, CA, USA, 87–94.

[15] Christopher Clarke, Doga Cavdir, Patrick Chiu, Laurent Denoue, and Don Kimber. 2020. Reactive Video: Adaptive Video Playback Based on User Motion for Supporting Physical Activity. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 196–208.

[16] Sebastian Cmentowski, Sukran Karaosmanoglu, Lennart E. Nacke, Frank Steinicke, and Jens Harald Krüger. 2023. Never Skip Leg Day Again: Training the Lower Body with Vertical Jumps in a Virtual Reality Exergame. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 786, 18 pages. https://doi.org/10.1145/3544548.3580973

[17] Alexandra Covaci, Anne-Hélène Olivier, and Franck Multon. 2015. Visual perspective and feedback guidance for vr free-throw training. *IEEE computer graphics and applications* 35, 5 (2015), 55–65.

[18] Alana Elza Fontes Da Gama, Thiago de Menezes Chaves, Pascal Fallavollita, Lucas Silva Figueiredo, and Veronica Teichrieb. 2019. Rehabilitation motion recognition based on the international biomechanical standards. *Expert Systems with Applications* 116 (2019), 396–409.

[19] Christof Faller. 2004. *Parametric coding of spatial audio.* Technical Report. EPFL.

[20] Jamie Ferguson, Lorna Paul, and Stephen Brewster. 2021. A Peripheral Tactile Feedback System for Lateral Epicondilytus Rehabilitation Exercise. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.

[21] Antonio Frisoli, Caterina Procopio, Carmelo Chisari, Ilaria Creatini, Luca Bonfiglio, Massimo Bergamasco, Bruno Rossi, and Maria Chiara Carboncini. 2012. Positive effects of robotic exoskeleton training of upper limb reaching movements after stroke. *Journal of neuroengineering and rehabilitation* 9, 1 (2012), 1–16.

[22] Nadia Garcia-Hernandez and Vicente Parra-Vega. 2009. Active and efficient motor skill learning method used in a haptic teleoperated system. In *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 915–920.

[23] Jeremy Raboff Gordon, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, Amos Miller, and Mar Gonzalez-Franco. 2023. Hearing the Way Forward: Exploring Ambient Navigational Awareness with Reduced Cognitive Load through Spatial Audio-AR. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 155, 7 pages. https://doi.org/10.1145/3544549.3585800

[24] Natsuki Hamanishi and Jun Rekimoto. 2020. PoseAsQuery: Full-Body Interface for Repeated Observation of a Person in a Video with Ambiguous Pose Indexes and Performed Poses. In *Proceedings of the Augmented Humans International Conference*. ACM, New York, NY, USA, 1–11.

[25] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. 2017. My Tai-Chi coaches: an augmented-learning tool for practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference*. ACM, New York, NY, USA, 1–4.

[26] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[27] WILLIAM HARTMANN and Eric Macaulay. 2014. Anatomical limits on interaural time differences: an ecological perspective. *Frontiers in Neuroscience* 8 (2014). https://doi.org/10.3389/fnins.2014.00034

[28] Thuong N Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. 2016. Onebody: remote posture guidance system using first person view in virtual environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction*. ACM, New York, NY, USA, 1–10.

[29] Eve Hoggan, Andrew Crossan, Stephen A Brewster, and Topi Kaaresoja. 2009. Audio or tactile feedback: which modality when?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, New York, NY, USA, 2253–2256.

[30] Xuhui Hu, Aiguo Song, Zhikai Wei, and Hong Zeng. 2022. StereoPilot: A wearable target location system for blind and visually impaired using spatial audio rendering. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30 (2022), 1621–1630.

[31] Xuhui Hu, Aiguo Song, Hong Zeng, and Dapeng Chen. 2022. Intuitive environmental perception assistance for blind amputees using spatial audio rendering. *IEEE Transactions on Medical Robotics and Bionics* 4, 1 (2022), 274–284.

[32] Masahiko Inami, Daisuke Uriu, Zendai Kashino, Shigeo Yoshida, Hiroto Saito, Azumi Maekawa, and Michiteru Kitazaki. 2022. Cyborgs, Human Augmentation, Cybernetics, and JIZAI Body. In *Proceedings of the Augmented Humans International Conference 2022* (Kashiwa, Chiba, Japan) *(AHs '22)*. Association for Computing Machinery, New York, NY, USA, 230–242. https://doi.org/10.1145/3519391.3519401

[33] Md Shafiqul Islam and Sol Lim. 2022. Vibrotactile feedback in virtual motor learning: A systematic review. *Applied Ergonomics* 101 (2022), 103694.

[34] Norihide Itoh, Hitoshi Kagaya, Eiichi Saitoh, Kei Ohtsuka, Junya Yamada, Hiroki Tanikawa, Shigeo Tanabe, Naoki Itoh, Takemitsu Aoki, and Yoshikiyo Kanada. 2012. Quantitative assessment of circumduction, hip hiking, and forefoot contact gait using Lissajous figures. *Japanese Journal of Comprehensive Rehabilitation Science* 3 (2012), 78–84.

[35] Florian Jeanne, Yann Soullard, Ali Oker, and Indira Thouvenin. 2017. EBAGG: Error-based assistance for gesture guidance in virtual environments. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*. IEEE, Long Beach, CA, USA, 472–476.

[36] Florian Jeanne, Indira Thouvenin, and Alban Lenglet. 2017. A study on improving performance in gesture training through visual guidance based on learners' errors. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, New York, NY, USA, 1–10.

[37] Daniel Johnston, Hauke Egermann, and Gavin Kearney. 2019. Measuring the behavioral response to spatial audio within a multi-modal virtual reality environment in children with autism spectrum disorder. *Applied Sciences* 9, 15 (2019), 3152.

[38] Jakob Karolus, Felix Bachmann, Thomas Kosch, Albrecht Schmidt, and Paweł W. Woźniak. 2021. Facilitating Bodily Insights Using Electromyography-Based Biofeedback during Physical Activity. In *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction* (Toulouse & Virtual, France) *(MobileHCI '21)*. Association for Computing Machinery, New York, NY, USA, Article 14, 15 pages. https://doi.org/10.1145/3447526.3472027

[39] Rushil Khurana, Karan Ahuja, Zac Yu, Jennifer Mankoff, Chris Harrison, and Mayank Goel. 2018. GymCam: Detecting, recognizing and tracking simultaneous exercises in unconstrained scenes. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (2018), 1–17.

[40] Yousef Kowsar, Masud Moshtaghi, Eduardo Velloso, Lars Kulik, and Christopher Leckie. 2016. Detecting unseen anomalies in weight training exercises. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. ACM, New York, NY, USA, 517–526.

[41] Kartikaeya Kumar, Lev Poretski, Jiannan Li, and Anthony Tang. 2022. Tourgether360: Collaborative Exploration of 360° Videos Using Pseudo-Spatial Navigation. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 546 (nov 2022), 27 pages. https://doi.org/10.1145/3555604

[42] Sih-Pin Lai, Cheng-An Hsieh, Yu-Hsin Lin, Teepob Harutaipree, Shih-Chin Lin, Yi-Hao Peng, Lung-Pan Cheng, and Mike Y Chen. 2020. StrengthGaming: Enabling dynamic repetition tempo in strength training-based exergame design. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–8.

[43] Marcel B Lanza, Thomas G Balshaw, and Jonathan P Folland. 2019. Is the joint-angle specificity of isometric resistance training real? And if so, does it have a neural basis? *European Journal of Applied Physiology* 119 (2019), 2465–2476.

[44] Jiann-Der Lee, Chung-Hung Hsieh, and Ting-Yang Lin. 2014. A Kinect-based Tai Chi exercises evaluation system for physical rehabilitation. In *2014 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, Long Beach, CA, USA, 177–178.

[45] Zhipeng Li, Yu Jiang, Yihao Zhu, Ruijia Chen, Ruolin Wang, Yuntao Wang, Yukang Yan, and Yuanchun Shi. 2022. Modeling the Noticeability of User-Avatar Movement Inconsistency for Sense of Body Ownership Intervention. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–26.

[46] Jeff Lieberman and Cynthia Breazeal. 2007. TIKL: Development of a wearable vibrotactile feedback suit for improved human motor learning. *IEEE Transactions on Robotics* 23, 5 (2007), 919–926.

[47] James Makous and John Middlebrooks. 1990. Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America* 87 (06 1990), 2188–200. https://doi.org/10.1121/1.399186

[48] Alexander Marquardt, Christina Trepkowski, Tom David Eibich, Jens Maiero, Ernst Kruijff, and Johannes Schöning. 2020. Comparing non-visual and visual guidance methods for narrow field of view augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics* 26, 12 (2020), 3389–3401.

[49] Troy L McDaniel, Morris Goldberg, Shantanu Bala, Bijan Fakhri, and Sethuraman Panchanathan. 2012. Vibrotactile feedback of motor performance errors for enhancing motor learning. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, New York, NY, USA, 419–428.

[50] Catarina Mendonça and Victoria Korshunova. 2020. Surround sound spreads visual attention and increases cognitive effort in immersive media reproductions. In *Proceedings of the 15th International Audio Mostly Conference*. 16–21.

[51] Microsoft. 2023. Microsoft Soundscape. https://www.microsoft.com/en-us/research/product/soundscape/.

[52] John Middlebrooks and David Green. 1991. Sound Localization by Human Listeners. *Annual review of psychology* 42 (02 1991), 135–59. https://doi.org/10.1146/annurev.ps.42.020191.001031

[53] Allen William Mills. 1958. On the minimum audible angle. *The Journal of the Acoustical Society of America* 30, 4 (1958), 237–246.

[54] Vimal Mollyn, Riku Arakawa, Mayank Goel, Chris Harrison, and Karan Ahuja. 2023. IMUPoser: Full-Body Pose Estimation Using IMUs in Phones, Watches, and Earbuds. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 529, 12 pages. https://doi.org/10.1145/3544548.3581392

[55] Arinobu Niijima, Toki Takeda, Kentaro Tanaka, Ryosuke Aoki, and Yukio Koike. 2021. Reducing muscle activity when playing tremolo by using electrical muscle stimulation to learn efficient motor skills. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (2021), 1–17.

[56] Ryogo Niwa, Kazuya Izumi, Shieru Suzuki, and Yoichi Ochiai. 2022. EMS-Supported Throwing: Preliminary Investigation on EMS-Supported Training of Movement Form. In *Universal Access in Human-Computer Interaction. Novel Design Approaches and Technologies: 16th International Conference, UAHCI 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part I*. Springer, 459–476.

[57] Eva Ostertagova, Oskar Ostertag, and Jozef Kováč. 2014. Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials* 611 (08 2014), 115–120.

[58] Max Pfeiffer, Tim Duente, and Michael Rohs. 2016. Let your body move: a prototyping toolkit for wearable force feedback with electrical muscle stimulation. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 418–427.

[59] Kun Qian, Chenshu Wu, Zimu Zhou, Yue Zheng, Zheng Yang, and Yunhao Liu. 2017. Inferring motion direction using commodity wi-fi for interactive exergames. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 1961–1972.

[60] Fazlay Rabbi, Taiwoo Park, Biyi Fang, Mi Zhang, and Youngki Lee. 2018. When virtual reality meets internet of things in the gym: Enabling immersive interactive machine exercises. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 2, 2 (2018), 1–21.

[61] T. Reilly. 2009. *Ergonomics in Sport and Physical Activity.* Human Kinetics. https://books.google.co.jp/books?id=rPJ6DwAAQBAJ
[62] V.J.B. Rice. 1998. *Ergonomics in Health Care and Rehabilitation.* Elsevier Health Sciences. https://books.google.co.jp/books?id=jDxtAAAAMAAJ
[63] Brad J Schoenfeld and Jozo Grgic. 2020. Effects of range of motion on muscle development during resistance training interventions: A systematic review. *SAGE open medicine* 8 (2020), 2050312120901559.
[64] Christian Schönauer, Kenichiro Fukushi, Alex Olwal, Hannes Kaufmann, and Ramesh Raskar. 2012. Multimodal motion guidance: techniques for adaptive and dynamic feedback. In *Proceedings of the 14th ACM international conference on Multimodal interaction.* ACM, New York, NY, USA, 133–140.
[65] Konstantin Semionov and Iain McGregor. 2020. Effect of various spatial auditory cues on the perception of threat in a first-person shooter video game. In *Proceedings of the 15th International Audio Mostly Conference.* 22–29.
[66] Seven. 2020. Seven. https://seven.app/.
[67] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. 2013. Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review* 20, 1 (2013), 21–53.
[68] Alexa F Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. 2020. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In *Proceedings of the 2020 CHI conference on human factors in computing systems.* 1–13.
[69] Mel Slater, Bernhard Spanlang, Maria V Sanchez-Vives, and Olaf Blanke. 2010. First person experience of body transfer in virtual reality. *PloS one* 5, 5 (2010), e10564.
[70] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 179–188.
[71] Maurício Sousa, João Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. 2016. SleeveAR: Augmented reality for rehabilitation using realtime feedback. In *Proceedings of the 21st international conference on intelligent user interfaces.* ACM, New York, NY, USA, 175–185.
[72] Gabriele Spina, Guannan Huang, Anouk Vaes, Martijn Spruit, and Oliver Amft. 2013. COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing.* ACM, New York, NY, USA, 597–606.
[73] Thomas Stütz, Michael Domhardt, Gerlinde Emsenhuber, Daniela Huber, Martin Tiefengrabner, Nicholas Matis, and Simon Ginzinger. 2017. An interactive 3D health app with multimodal information representation for frozen shoulder. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services.* ACM, New York, NY, USA, 1–11.
[74] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@ Home: Exploring visual guidance and feedback techniques for physiotherapy exercises. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 4123–4132.
[75] Yoshimasa Tokuyama, RPC Janaka Rajapakse, Sachiyo Yamabe, Kouichi Konno, and Yi-Ping Hung. 2019. A Kinect-Based Augmented Reality Game for Lower Limb Exercise. In *2019 International Conference on Cyberworlds (CW).* IEEE, Long Beach, CA, USA, 399–402.
[76] Milka Trajkova and Francesco Cafaro. 2018. Takes Tutu to ballet: designing visual and verbal feedback for augmented mirrors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–30.
[77] Hsin-Ruey Tsai, Yuan-Chia Chang, Tzu-Yun Wei, Chih-An Tsao, Xander Chin-yuan Koo, Hao-Chuan Wang, and Bing-Yu Chen. 2021. GuideBand: Intuitive 3D Multilevel Force Guidance on a Wristband in Virtual Reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–13.
[78] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. Motionma: motion modelling and analysis by demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 1309–1318.
[79] H. Wallach. 1940. The Role of Head Movements and Vestibular and Visual Cues in Sound Localization. *Journal of Experimental Psychology* 27, 4 (1940), 339. https://doi.org/10.1037/h0054629
[80] Zhu Wang, Anat Lubetzky, Marta Gospodarek, Makan TaghaviDilamani, and Ken Perlin. 2019. Virtual environments for rehabilitation of postural control dysfunction. *arXiv preprint arXiv:1902.10223* (2019).
[81] Lauren Ward, Ben G Shirley, et al. 2019. Personalization in object-based audio for accessibility: A review of advancements for hearing impaired listeners. *Journal of the Audio Engineering Society* 67, 7/8 (2019), 584–597.
[82] Robert Sessions Woodworth and Harold Schlosberg. 1954. *Experimental psychology.* Oxford and IBH Publishing.
[83] Chengshuo Xia, Xinrui Fang, Riku Arakawa, and Yuta Sugiura. 2022. VoLearn: A Cross-Modal Operable Motion-Learning System Combined with Virtual Avatar and Auditory Feedback. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–26.

[84] Ungyeon Yang and Gerard Jounghyun Kim. 2002. Implementation and evaluation of "just follow me": An immersive, VR-based, motion-training system. *Presence: Teleoperators & Virtual Environments* 11, 3 (2002), 304–323.
[85] Zhijian Yang, Yu-Lin Wei, Sheng Shen, and Romit Roy Choudhury. 2020. Ear-ar: indoor acoustic augmented reality on earphones. In *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*. 1–14.
[86] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. 2020. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. ACM, New York, NY, USA, 1–14.
[87] Huiyu Zhou and Huosheng Hu. 2008. Human motion tracking for rehabilitation—A survey. *Biomedical signal processing and control* 3, 1 (2008), 1–18.
[88] Franz Zotter and Matthias Frank. 2019. *Ambisonics: A practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality*. Springer Nature.

Received February 2024; revised May 2024; accepted June 2024