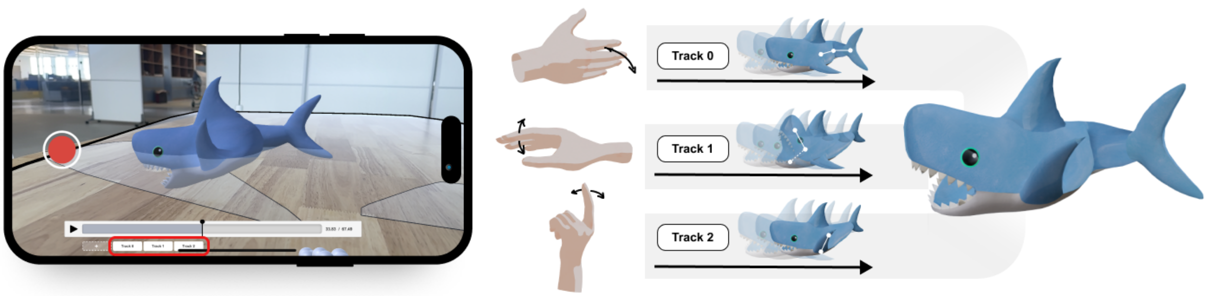
AgileFingers: Authoring AR Character Animation Through Hierarchical and Embodied Hand Gestures

Yue Lin*
Computational Media and Arts
HKUST(GZ)

Yudong Huang
Computational Media and Arts
HKUST(GZ)

David Yip
Computational Media and Arts
HKUST(GZ)

Zeyu Wang
Computational Media and Arts
HKUST(GZ)



(a) The overall user interface of the animating system

(b) Hierarchical control to compose a shark swinging, opening mouth and waving fin

Figure 1: The AgileFingers system allows users to (a) author character animation on mobile AR, (b) use hand gestures to embody the character through hierarchical control and composite the local movements together

Abstract

We present AgileFingers, a hand gesture-based solution for authoring AR character animation based on a mobile device. Our work initially categorizes four major types of animals under Vertebrata. We conducted a formative study on how users construct hierarchical relationships in full-body skeletal animation and potential hand structure mapping rules. Informed by the study, we developed a hierarchical segmented control system, which enables novice users to manipulate full-body 3D characters with unimanual gestures sequentially. Our user study reveals the ease of use, intuitiveness, and high adaptability of the AgileFingers system across various characters.

Index Terms: Human-centered computing—Human Computer interaction (HCI)—Interaction paradigms—Mixed / augmented Reality; Human-centered computing—Human Computer interaction (HCI)—Interaction techniques—Gestural input

1 Introduction

3D character animation is actively involved in extensive applications across many mobile platforms for its capability to enhance users' immersive experience concretely [2]. However, the animation creation precludes non-professional users from customization on mobile platforms due to the heavy load of traditional keyframing editing systems.

Addressing the lack of intuitiveness inherent in the keyframing system, motion capture (MoCap) animates the characters in a performance-based manner. However, current MoCap solutions predominantly transform full-body movements into characters. It is evident that capturing entire body motions with handheld mobile devices or even headsets is not feasible due to spatial constraints. Inspired by the expressiveness of sign language, we leverage the dexterity of hands for embodying 3D character animations.

Animating characters with hand movements, also recognized as hand-based digital puppetry, involves mapping the movements of hand joints to specific skeleton joints. Existing hand gesture-based 3D animation methods typically control the entire model's motion simultaneously [1]. This one-to-one mapping, however, is only effective when the model's skeleton nodes do not exceed the number of hand joints. Moreover, the human hand's anatomical structure inherently limits the ability to depict certain actions, especially those involving less dexterous fingers. To surmount these limitations and enable more versatile model control, one possible way is to segment the character skeleton and animate the pieces sequentially, which has been well-established in full-body motion capture [3]. However, there's a noticeable gap in research on a hierarchical skeletal control framework tailored to the hand's anatomy. Therefore, we propose a sequential, step-by-step control of character motion. To satisfy user's need in segmenting the character, the critical question arises: which skeletal structures are fundamental in character topology? How do users comprehend these structures towards hands?

To address this, we studied skeletal types across various species and conducted a user study to devise a taxonomy of fundamental skeletal structures and associated constraint rules for hand mapping strategy. Building on these findings, we introduce AgileFingers, a system aimed at validating the efficacy of swift animation prototyping using this hierarchical control approach. AgileFingers consists of two core modules: the **mapping** system enables users to customize the mapping between the hand and the character's skeleton under the established constraint rules, while the **animation** system offers users the flexibility to animate the character based on their unique mapping configuration.

2 The AgileFingers System

2.1 Design Considerations

We aim to design a universal hand-controlling scheme, ensuring hand movements can competently depict most 3D character topologies. To achieve this, we investigated the potential skeleton topologies that exist in character design and aimed to find users' controlling preferences for them.

We went through the animals under Chordata - Vertebrata based on the Linnaean Taxonomy, aggregating 22 animal orders with 48 representative species under 32 suborders. We further categorized the 48 species based on the locomotion pattern and limb differ-

*e-mail: ylin491@connect.hkust-edu.cn

(a) Mapping stage (b) Animation stage (c) Composite animation preview

Figure 2: The complete flow of animating a shark in the AgileFingers system. (a) The mapping stage of the system. Track 0 controls the spine, and track 1 controls the left fin of the shark, as indicated on the left. (b) The animation stage. The user can click on the track below the timeline to switch the tracks. (c) Preview the composite animation from all angles in AR.

entiation subjectively, and forms a taxonomy of 4 major classes: Limb-free, Wings, Arms, and Legs.

We further invited 12 participants to use their hands to mimic the animation of these 4 classes of characters to understand: (1) What is the mapping strategy between hand structure and character topology? (2) What is the animation strategy for a complex movement, and how will the participants decompose it?

The results show that although each person's **mapping strategy** is different, the binding strategy embodies a high degree of personal individual specificity. In terms of the finger choices, not all fingers are used in the mapping because participants may use them for comfort or to help them comprehend the character structure. This indicates that in system design, it is necessary to give the mapping rights to the user and allow leaving finger joints to blank.

Users show two **animation decomposition strategies**: Starting from dominant moving part to limbs; Starting from global movement to local. This indicates that in the system design, the global movement of the character can be given priority, and then the sub-level actions can be stacked in turn.

2.2 Mapping Stage

AgileFingers operates through two phases: Mapping and Animating. The demo is implemented on iOS utilizing Unity, with hand capture facilitated by MediaPipe.

First, the user defines the mapping solutions for each recording track. Users could indicate the correspondence between their hand joints and the character skeleton highlighted in the semi-transparent model displayed in Augmented Reality (AR). This step establishes a correspondence between the finger joints and the skeleton nodes by engaging in a point-and-select mechanism, progressively assembling the selecting pairs onto the animation recording track. A character's most granular controllable unit is an individual skeleton, predominantly structured as chains. Users are tasked with selecting skeletons that mirror the hand's architecture - identical in the number of nodes and their hierarchical relationships. An exception arises with Y-shaped skeletal structures, such as a shark's mouth, where multiple subsidiary bones are concurrently linked to a singular parent bone. Here, one bone is designated as the primary control bone. The focus is solely on the movements of this primary control bone in relation to its parent during manipulation.

2.2.1 Animation Stage

In this stage, users sequentially record each track defined in the previous stage by performing in front of the camera. The base track has the root node that controls the global shift of the character. The sub-movements are overlaid layer by layer onto the previous motions. Real-time animation previews, including all previously recorded tracks, are available to users during this recording process. Upon completion, users can view the recorded animation from all perspectives by walking around in the space.

3 Discussion

AgileFingers utilizes hand gestures and AR to enable novice users to customize 3D character animation prototypes on mobile devices quickly. We conducted a user study on 12 participants. The results reveal the significant intuitiveness and entertainment of the system.

We acknowledge that the animation performance was not computationally optimized because we primarily focused on the system's feasibility in producing the character animation using a hierarchy of hand embodiment. The current system reveals great potential in controlling non-human avatars, especially with Limb-free characters and those in cartoon styles, but shows a deficit in humanoid characters and those that have chain structures with more than 4 nodes. This could be possibly alleviated by introducing inverse kinematics under the current scheme to reduce the redundant control in non-endpoints skeleton nodes.

AgileFingers adopts point-and-click in mapping creation since we want to ensure the users can understand the mapping relationship and maintain originality in designing the preferred controlling scheme. The data could be used to instruct auto-mapping generation based on the user's intention or provide recommended mappings on common structures in future works.

The AgileFingers demo is carried out with the built-in camera and LiDAR sensor carried on iPhone 14 Pro and calculated by MediaPipe. This lightweight solution reveals the great potential of the AgileFingers system to migrate to mobile devices with higher hardware support or AR/VR headsets as either a creation tool or an entertainment application. We believe AgileFingers can provide a starting point to better involve novice users in enjoying 3D animation creation.

References

[1] Y. Jiang, Z. Li, M. He, D. Lindlbauer, and Y. Yan. HandAvatar: Embodying Non-Humanoid Virtual Avatars through Hands. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581027

[2] Y. Rekik, L. Grisoni, and N. Roussel. Towards many gestures to one command: A user study for tabletops. In *Human-Computer Interaction – INTERACT 2013*, pp. 246–263. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[3] Y. Seol, C. O'Sullivan, and J. Lee. Creature Features: Online Motion Puppetry for Non-Human Characters. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '13, p. 213–221. Association for Computing Machinery, New York, NY, USA, 2013. doi: 10.1145/2485895.2485903