Towards Multimodal Interaction with AI-Infused Shape-Changing Interfaces

Chenfeng Gao*
jessegao7@uchicago.edu
University of Chicago
Chicago, IL, USA

Wanli Qian*
michaelq@uchicago.edu
University of Chicago
Chicago, IL, USA

Richard Liu
guanzhi@uchicago.edu
University of Chicago
Chicago, IL, USA

Rana Hanocka
ranahanocka@uchicago.edu
University of Chicago
Chicago, IL, USA

Ken Nakagaki
knakagaki@uchicago.edu
University of Chicago
Chicago, IL, USA

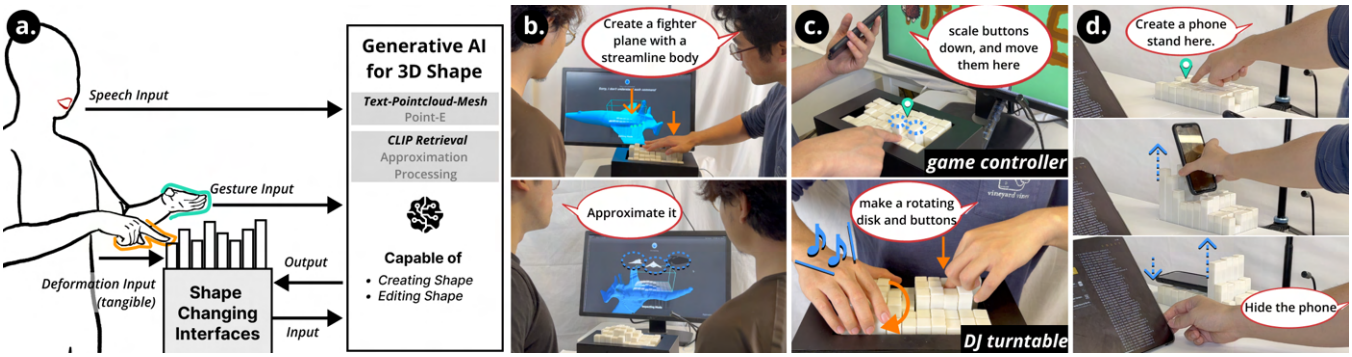

Figure 1: (a.) Interaction framework. Applications for (b.) 3D modeling, (c.) adaptive controller, (d.) reconfigurable tabletop.

ABSTRACT

We present a proof-of-concept system exploring multimodal interaction with AI-infused Shape-Changing Interfaces. Our prototype integrates inFORCE, a 10x5 pin-based shape display, with AI tools for 3D mesh generation and editing. Users can create and modify 3D shapes through speech, gesture, and tangible inputs. We demonstrate potential applications including AI-assisted 3D modeling, adaptive physical controllers, and dynamic furniture. Our implementation, which translates text to point clouds for physical rendering, reveals both the potential and challenges of combining AI with shape-changing interfaces. This work explores how AI can enhance tangible interaction with 3D information and opens up new possibilities for multimodal shape-changing UIs.

CCS CONCEPTS

• **Human-centered computing** → **Interaction devices**.

*Both authors contributed equally to this research.

UIST Adjunct '24, October 13–16, 2024, Pittsburgh, PA, USA

ACM ISBN 979-8-4007-0718-6/24/10
https://doi.org/10.1145/3672539.3686315

KEYWORDS

Shape-Changing Interface, Multimodal Interaction, Generative AI

ACM Reference Format:
Chenfeng Gao, Wanli Qian, Richard Liu, Rana Hanocka, and Ken Nakagaki. 2024. Towards Multimodal Interaction with AI-Infused Shape-Changing Interfaces. In *The 37th Annual ACM Symposium on User Interface Software and Technology (UIST Adjunct '24), October 13–16, 2024, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3672539.3686315

1 INTRODUCTION

Researchers have long envisioned dynamic physical materials that can shape-shift and adapt according to users' intent, expressed via speech, tangible, and gestural interaction [8, 10]. Shape-Changing Interfaces (SCIs) have been one of the primary areas of research in designing such interaction by engineering actuated hardware systems [1, 5, 20]. However, enhancing the hardware capability to adapt to user intent dynamically requires software development in multimodal interaction (to understand user intent richly) and 3D shape generation (to render diverse shapes flexibly) [2, 3].

Recent advances in Generative Artificial Intelligence (Gen-AI) have opened new avenues for content creation and editing, particularly in text, image, and 3D shape generation [12, 16, 17]. This burgeoning field presents a vast opportunity for the SCIs to move closer to realizing our long-standing visions (e.g. Radical Atoms [10] or Claytronics [6]).

Towards this end, this paper introduces a proof-of-concept prototype that integrates AI-based 3D content generation with SCIs.

Our system explores multimodal interactions (speech, gesture, and touch) [7, 21] to enable the intuitive creation and manipulation of 3D shapes, physically rendered with an SCI. This approach uniquely combines Gen-AI with SCIs, opening up new possibilities of "AI-infused Shape-Changing UIs" facilitated by multimodal interactions.

While we previously explored the use of large language models (LLMs) for SCIs based on code-generation capability [18], this paper explores novel opportunities with AI-infused SCIs through 1. a proof-of-concept prototype fusing 3D shape generation AI and shape display hardware [13], and 2. exploration of potential applications.

2 IMPLEMENTATION AND INTERACTION

Our proof-of-concept system consists of a 10x5 pin-based shape display, inFORCE [13], an OAK-D depth camera for gesture recognition, and a microphone for speech input. Our prototype included a monitor as well for visual display of generated 3D models. Overall, the software processes user multimodal inputs and generates 3D meshes, displays them physically through the Shape display, and allows users to manipulate them accordingly ([A] in our video). For input processing, wit.ai [24] was used for *speech inputs*, MediaPipe [11] with customized API from depth camera for 3D *gesture inputs*, and inFORCE pin-displacement detection for *tangible inputs*. These inputs are translated into system commands using a Keyword Extraction Algorithm and Spacy [9] to map them to predefined command templates ([B] in our video).

We utilize OpenAI's Point-E [15] for text-to-mesh generation and a novel Mesh-to-Mesh retrieval algorithm based on CLIP [19] embeddings. Our 3D model dataset, combining COSEG [23], ShapeNetCore [4], and ShapeNetSem, totals 65,000 models with pre-computed latent embeddings for efficient retrieval. Mesh deformations are managed through tangible interactions, while affine transformations are handled through gestural inputs, allowing users to manipulate and modify the generated shapes directly. A virtual workspace developed in Unity provides a digital representation of the setup. It renders mesh results on the shape display and offers a GUI with visual feedback on speech recognition, AI mesh generation progress, and physical-virtual alignment.

We designed three interaction modes (Figure 2): CREATE generates shapes via speech and gestures; EDIT enables tangible manipulation, gesture-based pulling [2], and AI-suggested modifications/approximation; INSPECT allows tactile examination and multimodal position/orientation adjustments.

3 POTENTIAL APPLICATIONS

3D Modeling: Our system can enhance ideation, brainstorming, and customization in 3D design processes (Figure 1b). With AI-infused SCI, users can create 3D objects via speech commands and hand gestures, inspect and manipulate them using gestures, and edit them through tangible and gestural interactions. The AI approximation command can provide suggested 3D objects based on modified meshes, aiding the ideation process in a highly tangible manner. This setup may be useful in collaborative design scenarios, as envisioned in the Claytronics project [6]. Multiple users can generate shapes, physically inspect them, and make micro-adjustments collaboratively.

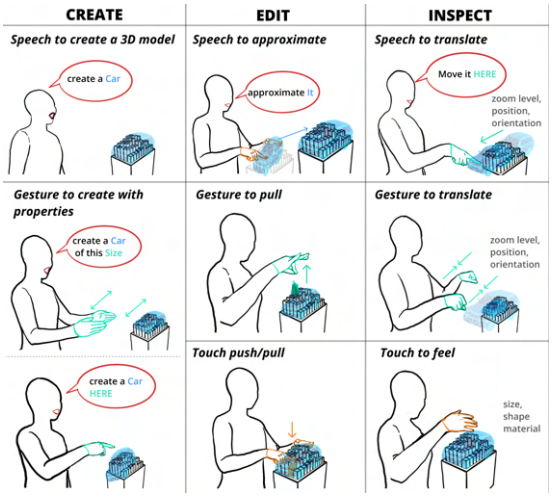

Figure 2: Interaction Design

Adaptive Controller: The system could allow users to create or summon customizable physical controllers of various sizes and shapes (Figure 1c.). Users can request specific controller layouts, such as large round buttons for palm interaction or smaller controls for single-hand operation. The system's flexibility allows for on-the-fly adjustments to suit changing needs during use. Musicians or DJs could benefit from this by creating customizable musical instrument interfaces on the shape display.

Adaptive Tabletop Furniture: Lastly, inspired by TRANSFORM [22], AI-infused SCI could create and control tabletop surface shapes through multimodal interaction (Figure 1d.). Users can request specific shapes for various purposes, such as a phone stand for hands-free calls or a wall divider to hide distractions [14]. The system can also generate and adjust objects like pen holders based on user needs.

4 CONCLUSION AND FUTURE WORK

This paper introduced a novel approach integrating Gen-AI with SCIs facilitated by multimodal interaction. Our prototype demonstrates the potential of this approach while revealing limitations and opportunities for future research.

Current challenges include *slow response times* (~60 seconds), *low-resolution shape output*, and *limited interaction* (e.g. only pre-defined gestures). Critically, we also learned employing text-to-mesh generation [15] has a limitation in that it does not take into account the hardware constraint (pin-display resolution and hardware geometry). Hence, we found the use of LLM for code-generation to be more adaptable to generate shape, motion and interaction [18].

By addressing these challenges, AI-infused SCIs should bring us closer to the vision of adaptive materials that can listen, transform, and be manipulated tangibly, opening new possibilities for HCI.

ACKNOWLEDGMENTS

We thank the CERES program and Prof. Andrew Chien at the University of Chicago for their support. We also appreciate the help of Actuated Experience Lab members with the project.

REFERENCES

[1] Jason Alexander, Anne Roudaut, Jürgen Steimle, Kasper Hornbæk, Miguel Bruns Alonso, Sean Follmer, and Timothy Merritt. 2018. Grand challenges in shape-changing interface research. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–14.

[2] Matthew Blackshaw, Anthony DeVincenzi, David Lakatos, Daniel Leithinger, and Hiroshi Ishii. 2011. Recompose: direct and gestural interaction with an actuated surface. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 1237–1242.

[3] Richard A Bolt. 1980. "Put-that-there" Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. 262–270.

[4] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. 2015. *ShapeNet: An Information-Rich 3D Model Repository*. Technical Report arXiv:1512.03012 [cs.GR]. Stanford University — Princeton University — Toyota Technological Institute at Chicago.

[5] Marcelo Coelho and Jamie Zigelbaum. 2011. Shape-changing interfaces. *Personal and Ubiquitous Computing* 15 (2011), 161–173.

[6] davidchiu21. 2006. *Claytronics - Physical Dynamic Rendering*. Retrieved March 31, 2023 from https://www.youtube.com/watch?v=bcaqzOUv2Ao

[7] Bruno Dumas, Denis Lalanne, and Sharon Oviatt. 2009. MultimodalMultimodal Interfaces: A Survey of Principles, Models and Frameworks. In *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 3–26. https://doi.org/10.1007/978-3-642-00437-7_1

[8] Seth C Goldstein and Todd C Mowry. 2004. Claytronics: A scalable basis for future robots. (2004).

[9] Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. (2017). To appear.

[10] Hiroshi Ishii, Dávid Lakatos, Leonardo Bonanni, and Jean-Baptiste Labrune. 2012. Radical atoms: beyond tangible bits, toward transformable materials. *interactions* 19, 1 (2012), 38–51.

[11] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. 2019. MediaPipe: A Framework for Building Perception Pipelines. arXiv:1906.08172 [cs.DC]

[12] Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, and Rana Hanocka. 2021. Text2Mesh: Text-Driven Neural Stylization for Meshes. arXiv:2112.03221 [cs.CV]

[13] Ken Nakagaki, Daniel Fitzgerald, Zhiyao (John) Ma, Luke Vink, Daniel Levine, and Hiroshi Ishii. 2019. InFORCE: Bi-Directional 'Force' Shape Display for Haptic Interaction. In *Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Tempe, Arizona, USA) *(TEI '19)*. Association for Computing Machinery, New York, NY, USA, 615–623. https://doi.org/10.1145/3294109.3295621

[14] Ken Nakagaki, Jordan L Tappa, Yi Zheng, Jack Forman, Joanne Leong, Sven Koenig, and Hiroshi Ishii. 2022. (Dis) Appearables: A Concept and Method for Actuated Tangible UIs to Appear and Disappear based on Stages. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–13.

[15] Alex Nichol, Heewoo Jun, Prafulla Dhariwal, Pamela Mishkin, and Mark Chen. 2022. Point-E: A System for Generating 3D Point Clouds from Complex Prompts. *arXiv preprint arXiv:2212.08751* (2022).

[16] OpenAI. 2023. *ChatGPT*. Retrieved April 5, 2023 from https://chat.openai.com/

[17] OpenAI. 2023. *DALL-E2*. Retrieved April 5, 2023 from https://openai.com/product/dall-e-2

[18] Wanli Qian, Chenfeng Gao, Anup Sathya, Ryo Suzuki, and Ken Nakagaki. 2024. SHAPE-IT: Exploring Text-to-Shape-Display for Generative Shape-Changing Behaviors with LLMs. In *The 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3654777.3676348

[19] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]

[20] Majken K Rasmussen, Esben W Pedersen, Marianne G Petersen, and Kasper Hornbæk. 2012. Shape-changing interfaces: a review of the design space and open research questions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 735–744.

[21] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern recognition letters* 36 (2014), 189–195.

[22] Luke Vink, Viirj Kan, Ken Nakagaki, Daniel Leithinger, Sean Follmer, Philipp Schoessler, Amit Zoran, and Hiroshi Ishii. 2015. Transform as adaptive and dynamic furniture. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 183–183.

[23] Xiao Wang, Jingen Liu, Tao Mei, and Jiebo Luo. 2021. CoSeg: Cognitively Inspired Unsupervised Generic Event Segmentation. arXiv:2109.15170 [cs.CV]

[24] wit.ai. 2023. *wit.ai*. https://wit.ai/