# Intuitive space texture generation using hand tracking, speech recognition, and generative AI

*Yudai* Watanabe[1,*] and *Michael* Cohen[1,**]

[1]Spatial Media Group, University of Aizu; Aizu-Wakamatsu, Fukushima; Japan

**Abstract.** This research aims to explore new methods of intuitively redesigning room interiors using gesture, speech, and generative AI. This approach represents a new approach to interior design, allowing users to easily customize appearance of a room through voice and hand gestures. This project investigates how hand tracking, speech recognition, and generative AI can be integrated to enable intuitive and user-friendly interior texture customization in virtual spaces. Previous studies on interior design using XR have mainly used augmented reality (AR) to relocate furniture. However, in these methods, the only way to select furniture textures is to search for them in prepared furniture. Our method uses hand-tracking and speech recognition to capture a user's desired image and employs generative AI to realize these preferences in a VR environment. The process involves scanning real-world furniture and rooms and applying AI-generated textures based on what the user communicates. The system allows users to easily visualize room interiors and modify them according to their preferences. This can enhance the traditional room design process. This method is currently restricted to texture only, but 3D model generation AI could provide additional flexibility. This method also has the potential for collaborative design work by sharing an environment.

## Keywords

hand tracking, gesture interpretation, speech recognition, image-generative AI, virtual reality, interior design

## 1 Introduction

Various effects of furniture tones and wall colors have been reported, such as improving mood and increasing work efficiency. However, when designing or redecorating a room, it is difficult to project one's mental image into the furniture. To solve this problem, there already exist applications that use Augmented Reality (AR) to virtually change furniture placement, wallpaper, and flooring. Diminished reality can virtually remove furniture from a real space in AR, making it seem as if it does not exist in that place. By using such technology, furniture can be rearranged. Applications that use Virtual Reality (VR) are highly immersive and can simulate room redecoration without being restricted to a real space. However, few of

*e-mail: donabe722@gmail.com
**e-mail: xilehence@gmail.com

them change the texture of existing furniture. We propose an intuitive method to change furniture textures using hand tracking, speech understanding, and image generation AI in VR space. One can intuitively create a room as imagined. Furthermore, it is possible to occupy a room styled with this system while immersed in VR, and it is also possible to show the simulated room to an interior designer to project it into actual furniture. Now that there are more inexpensive head-mounted displays (HMD) than ever before, this method will use emerging features and combine them with advanced image-generating AI to make it possible to generalize what was previously difficult to do without specialists.

## 2 Related Work

Sara et al. are working on capturing indoor scenes in 3D, detecting obstacles, mapping walkable areas, and generating virtual spaces [1]. The idea is that this will allow the use of real space even within immersive virtual spaces. This is similar to our research where a room is brought directly into the virtual space and its textures are changed. However, our research differs in that we import the type and size of furniture as well as the traversable area.

Siltanen et al. propose an interior design method using Augmented Reality and Diminished Reality [2]. The idea is to make already existing furniture appear as if it has been erased and to allow interior design using room structure and environmental lighting without distraction. This can allow for less constrained interior design compared to traditional AR-based furniture placement simulations. One of the many differences from our research is that the furniture placement is left up to the user. We are aiming for a more intuitive interior design.

Kaleja et al. developed a tool called Dynamic Real-Time Visualization (DRTV) for interior design that allows manipulation of objects, modification of surface characteristics, and creation of detailed budgets [3]. The study highlights the benefits of VR technology in interior design, including an intuitive design process, real-time visualization, and improved communication between designer and client. However, furniture and room data must be modeled and cannot be handled solely by the client. Our system focuses more on allowing the client to restyle the interior design alone.

## 3 Method

### 3.1 Development environment

**Our System**
HMD: Meta Quest 3
Game Engine: Unity 2021.3.23f1
Programming Language: C#
VR SDK: Meta XR All-in-One SDK
**Employed Services**
3D Scanner: Polycam
Speech recognition AI: Whisper
Prompt generation AI: ChatGPT
Image generation AI: Stable Diffusion
**Image-Generation Server Computer**
OS: Microsoft Windows 11 Home
CPU: AMD Ryzen 7 3700X 8-Core Processor
RAM: 32.0 GB
GPU: NVIDIA GeForce RTX 2060 SUPER

### 3.2 Acquisition of furniture data and construction of the virtual room

To measure the size of the room and to obtain the furniture arrangement and dimensions, we use an Apple iPad Pro equipped with a LiDAR sensor and the AR Kit, an AR API provided by Apple, and placement of furniture in Unity. We use an application called Polycam [4] to capture real spaces.

The captured data is not UV-unwrapped, so UV-unwrapping is required to apply the textures. The data is loaded into Blender, and smart UV unwrapping is applied to each piece of furniture. Next, the captured data with UV unwrapping applied is imported into Unity. Special materials are applied to furniture that can be retextured.

### 3.3 Controlling textures with speech

This system data flow is shown in Figure 1. This system was built with Unity and uses Meta Quest 3 as the VR HMD. The user wears the helmet, launches the system, then sees the room imported as described by the previous section. After selecting furniture in the method described following, the user speaks the image of the furniture. User speech data is converted to text using the Whisper API [5]. Voice data captured by Unity is in AudioClip format, but requires transcoding to WAVE or other audio formats to be used with Whisper.

Since the generated text is spoken language, it is not immediately suitable for direct use in image generation AI. Therefore, ChatGPT [6] is used to convert dictated text into a prompt appropriate for directing image generation AI. Speech recognition can understand Japanese, and ChatGPT returns English language prompts even when Japanese is originally spoken.

Example 1: “Make that desk a wooden design.” → “Texture, Desk, Wood”

The image-generative AI creates textures using Stable Diffusion text-to-image service from the prompts obtained and the selected furniture texture size. Stable Diffusion is accessed from the HMD by arranging a server for the Stable Diffusion WebUI [7] on a networked desktop PC.

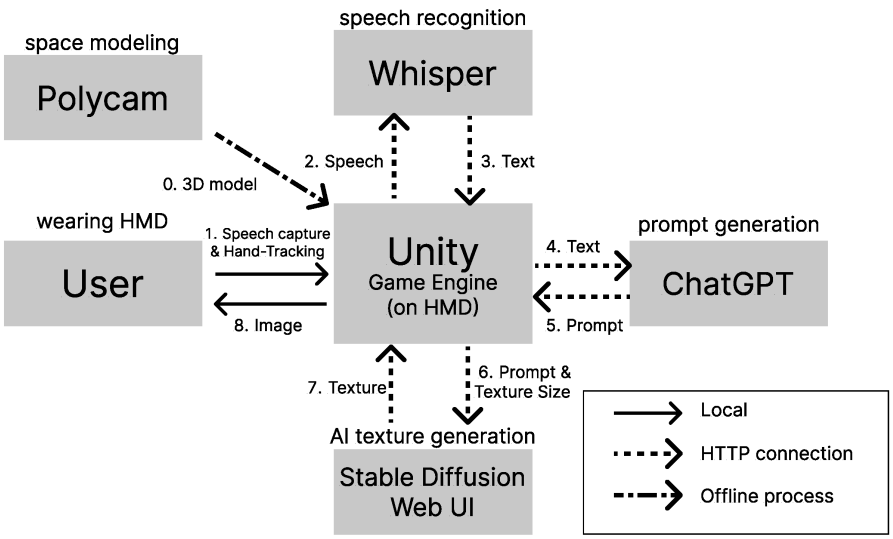
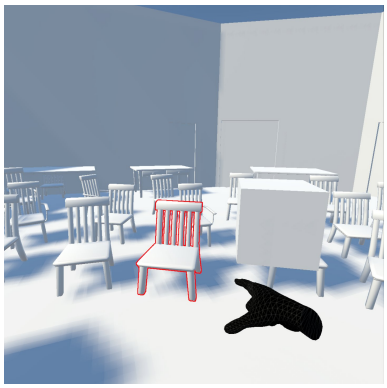


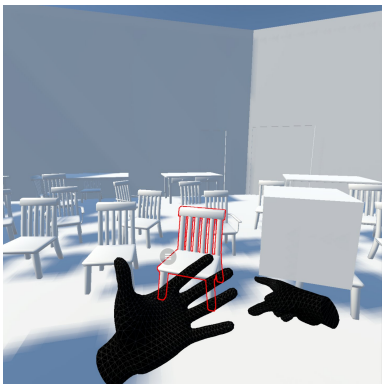

**Figure 1:** System Data Flow

### 3.4 User Interface

The user points to a specific piece of furniture to adjust its appearance, and the furniture is highlighted to confirm target acquisition. The user touches the thumb and index finger of the left hand to begin speech capture. After the user describes the desired image, the process executes online and adjusts the look of the furniture. (Figure 2)



**(a)** Point at a target to acquire



**(b)** Open left hand to start redesign


**(c)** Record speech for redesign furniture

**(d)** Change texture procedurally

**Figure 2:** Procedure for adjusting furniture textures

## 4 Results

The system is aimed at enabling users to intuitively change the look and feel of a room's interior. We predict that users will be able to generate and visualize interior design appearance by simply verbalizing a description of the image they want. This will surpass the previous concept of interior design and room redecoration, allowing users to build rooms more exactly as desired.

## 5 Limitations

There are two possible problems. First, users must verbalize their images. If the user does not have a concrete and detailed mental image, results may not be as expected. On the other hand, there may be cases in which the user can formulate a concrete image from only a vague image and description. This is because image-generating AI can generate an image from any instructions.

Second, it is not easy to change the position of furniture. In interior design, it is necessary to rearrange the position of furniture. Currently, however, furniture positions cannot be changed because the room is scanned and then imported into the system as a 3D model. The following are possible solutions. First, only the shape of the room could be captured in the 3D scan of the room, and the furniture can be instantiated using a 3D model prepared in advance or by using a 3D generation AI. With this method, it would be possible to design a room by specifying the position and type of furniture. Furthermore, a method is proposed to track furniture in real-time by acquiring the room as a point cloud and tracking the 3D furniture model [8]. With this method, there is no need to rescan the room when the furniture layout changes.

## 6 Conclusion

Our method is more intuitive than previous methods and allows anyone to design a room freely. This will lead to the creation of spaces that are comfortable and have effects on people's minds. As a future step, we would like to generate 3D models of complex shapes and furniture. Generative AI has been advancing rapidly in recent years, and high-quality 3D model generation AIs such as "Shap-E" [9] and "DreamGaussian" [10] are appearing even as this paper is being written. By integrating them, our system will be able to create spaces more flexibly as desired. Furthermore, we would like to enable group collaboration work. This would be a new way for multiple people to discuss interior design, such as in an office or a couple's room. It will also be possible to create interiors that would be impossible in reality. For example, it will be possible to make a ceiling disappear to reveal the sky, or to add effects and animation to furniture.

## References

[1] M. Sra, S. Garrido-Jurado, C. Schmandt, P. Maes, *Procedurally generated virtual reality from 3D reconstructed physical space*, in *VRST: Proc. ACM Conf. on Virtual Reality Software and Technology* (2016), pp. 191–200, ISBN 9781450344913, DOI 10.1145/2993369.2993372, https://doi.org/10.1145/2993369.2993372

[2] S. Siltanen, H. Saraspää, J. Karvonen, *A complete interior design solution with diminished reality*, in *ISMAR: Proc. IEEE Int. Symp. on Mixed and Augmented Reality* (2014), pp. 371–372, DOI 10.1109/ISMAR.2014.6948494

[3] P. Kaleja, M. Kozlovská, J. of Civil Engineering **12**, 109 (2017), DOI 10.1515/sspjce-2017-0011

[4] *Polycam — LiDAR & 3D Scanner for iPhone & Android*, https://poly.cam, (Accessed 07/16/2023)

[5] *Introducing ChatGPT and Whisper APIs*, https://openai.com/blog/introducing-chatgpt-and-whisper-apis, (Accessed 11/17/2023)

[6] *ChatGPT*, https://chat.openai.com, (Accessed 11/17/2023)

[7] *AUTOMATIC1111/stable-diffusion-webui: Stable Diffusion web UI*, https://github.com/AUTOMATIC1111/stable-diffusion-webui, (Accessed 11/17/2023)

[8] D. Lindlbauer, A.D. Wilson, *Remixed Reality: Manipulating Space and Time in Augmented Reality*, in *CHI Conf. on Human Factors in Computing Systems* (2018), pp. 1–13, ISBN 9781450356206, DOI 10.1145/3173574.3173703

[9] H. Jun, A. Nichol, *Shap-E: Generating Conditional 3D Implicit Functions*, in *arXiv preprint arXiv:2305.02463* (2023), DOI 10.48550/arXiv.2305.02463

[10] J. Tang, J. Ren, H. Zhou, Z. Liu, G. Zeng, *DreamGaussian: Generative Gaussian Splatting for Efficient 3D Content Creation*, in *arXiv preprint arXiv:2309.16653* (2023), DOI 10.48550/arXiv.2309.16653