Making-A-Scene: A Preliminary Case Study on Speech-based 3D Shape Exploration through Scene Modeling

Shantanu Vyas *
Ting-Ju Chen†
Ronak R. Mohanty ‡
J. Mike Walker '66 Department of
Mechanical Engineering
Texas A&M University
College Station, Texas 77843, USA

Vinayak R. Krishnamurthy
J. Mike Walker '66 Department of
Mechanical Engineering
and Department of Computer Science
and Engineering (by Affiliation)
Texas A&M University
College Station, Texas 77843, USA

We explore verbalization as a means for quick-and-dirty 3D shape exploration in early design. Our work stems from the knowledge gap that the fundamental principles necessary to operationalize speech as a viable means for describing and communicating 3D digital forms do not currently exist. To address this gap, we present a case study on 3D scene modeling within the context of interior design. For this, we implemented a constrained workflow wherein a user can iteratively generate variations of seed templates of objects in a scene through verbal input. Using this workflow as an experimental setup, we systematically study four aspects of speech-based shape exploration, namely, (a) design-in-context (creating one shape with respect or in relation to the other), (b) order independence (sequence of parts preferred in speech-based shape exploration), (c) multi-scale exploration (study how speech allows overview-then-detail modifications), and (d) semantic regions of interest (effectiveness of speech for modifying regions of a given object). We finally present an observational study with 6 participants selected from diverse backgrounds to better understand shape verbalization.

1 INTRODUCTION

Advances in computer graphics and interactive techniques have surely stuck to this dictum in enabling, facilitating, and promoting visual thinking [1] within the purview of industrial, product, and architectural design. There is extensive work on computer support for shape creation using direct manipulation, sketching, gestures, hand-held controllers, etc. In this work, we aim to investigate verbalization as a means to describe visual forms (geometry) in the exploratory early stages of design.

Our work is motivated from the observation that early stage design exploration involves multi-modal thinking and communication; wherein the visual and verbal modes of thinking and communication offer *diverse and supportive roles* for each other [2]. In fact, verbal communication is especially central to early design when the necessity to generate new ideas quickly is greater than specifying on one single idea in detail [3, 4, 5]. Despite this, there are no systematic studies on verbal modalities for form exploration *in isolation*. Much of the work done on verbal communication is focused on capturing, visualizing, and summarizing conversations. Our broader goal is to understand how speech could be meaningfully utilized in future digital systems for iterative shape exploration.

*Email: svyas@tamu.edu
†Email: carol0712@tamu.edu
‡Email: ronakmohanty@tamu.edu
Email: vinayak@tamu.edu, Address all correspondence to this author.

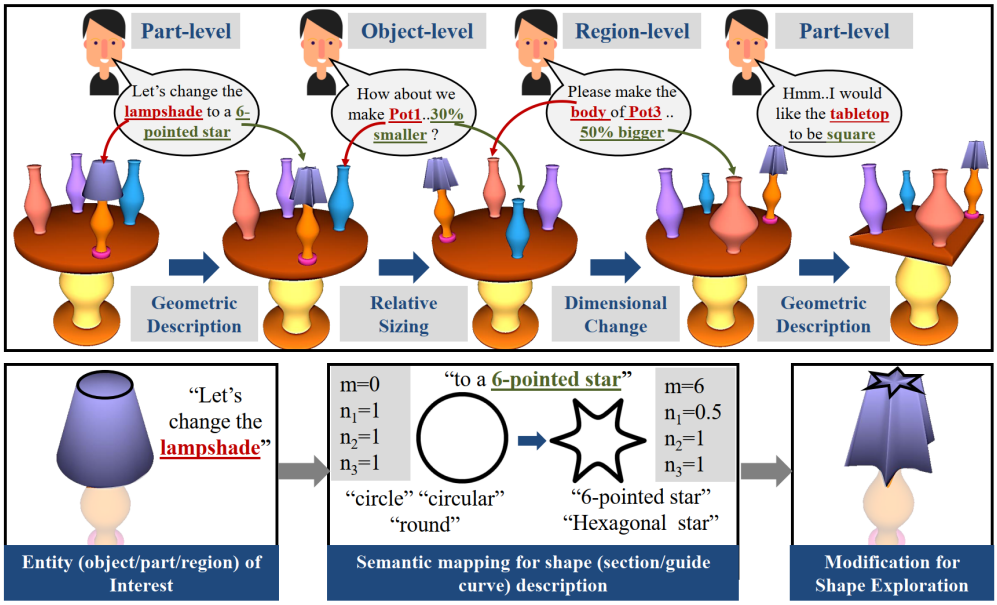

Fig. 1. The speech-assisted workflow in our case study (top) allows a user to start with a template scene (consisting of three pots and a lamp on top of a table) and iteratively edit this scene (shapes and sizes of objects) through verbal input. The shape modifications are powered by a combination of lofts and Superformulae (bottom).

1.1 Challenges & Need

The use of language in early design can be ambiguous and vague, which, on one hand fosters design exploration [6], but on the other, produces major hurdles for computational support tools that are currently inept in understanding and supporting informal and flexible design conversations [7]. As a result, prior works that integrate speech in design tools take a constrained approach and fall under the category of: (a) multimodal interfaces (e.g., gestures + speech) [8, 9, 10, 11, 12]; and/or (b) integrating voice-commands to existing CAD software [13, 14]. Prior works that use multimodal interactions include the early and seminal work by Bolt [8], *"Put-That-There"*, that combines voice inputs with gestures to create and manipulate (translate, rotate, etc.) 2D primitive shapes on a screen. Similarly, Gao et al., [10] use voice commands in a VR-environment, primarily to instantiate primitive shapes (cubes, cylinder, cones) and provide rough locations, with majority of the shape manipulation done through a 3D mouse. Recently, the work by Nanjundaswamy et al., [12] integrates gestures, brain-computer interface and speech in a CAD software, with the role of speech being limited to creating only circles, rectangles, orbits and arcs. The works by Sharma et al. [14], and Kou et al. [13], also extend on existing CAD software by providing speech and gesture input as an alternative to the software's existing features, which inherently do not support quick exploration in early design.

While there have been prior works that study verbalization in the design process, they are typically limited to specific scenarios and have not yet been implemented into computational support tools. For instance, Wiegers et al. [15] study how people describe shapes and their operations, by using ten pairs of pictures of arbitrary clay models and a few products to initiate the shape descriptions. Khan et al., [16, 17, 18] have studied the use of speech and gestures by architects and engineers limited to CAD-specific functions and procedures (such as rotation, copy, move, extrude etc.). More recently, the work by Ungureanu et al., [6] focuses on understanding frequently used natural language expressions in design conversations between architects, where they show how ambiguity and vagueness is prevalent in early design and reduces at later stages of the design process. However, there is a need for a more constrained study of how people describe shapes (in the absence of different modalities) for quick design ideation, which can also be operationalized to a certain degree. The prior works show that a constrained and concentrated effort is needed to accomplish this, before speech can be used in a completely natural and generalized form for shape exploration.

1.2 Approach & Rationale

Our goal, in this work, is not to design a full-fledged feature rich speech-based design system. Instead, our strategy is to perform a focused study of four aspects of speech-based shape exploration, namely:

1. *Design-in-context:* Study exploration of different shapes in the context of a given scene.
2. *Order Independence:* Study the sequence of user-preferred parts and objects.
3. *Multi-scale exploration:* Observe users' utilization of a hierarchical (overview-then-detail) approach to shape modification through speech.
4. *Semantic Regions-of-interest:* Study effectiveness of speech in modifying multi-region shapes.

1.3 Contribution

To study these four aspects, we develop a constrained 3D interior scene modeling application (Figure 1) that comprises of seed shape templates of known objects (lamps, table, pots). The purpose for a scene is to facilitate contextual thinking in relation to a meaningful design problem. Using 3D lofts (generalized cylinders) as our shape representation, this system enables users to iteratively edit and explore the template scene using their speech. Using this system as an experimental setup, we conduct a case study with six individuals specifically selected from different design backgrounds (engineering, architecture, visualization). We present our findings on the patterns of user behavior, and their actions with respect to the semantic representation of shapes.

2 RELATED WORKS

2.1 Speech-Based Workflows in Design

In design, sketching has been considered as a main communication method between designers for long. However, a comparison study conducted by Jonson [4] found that verbalization is the primary tool for getting started on a design and externalizing the "Aha!" moment. In fact, researchers have been devoted towards studying different kinds of verbalization to demonstrate its efficacy in design problem solving and as an aid to the designer's thinking process [19, 20]. Several works by Adler [21, 22] also emphasize upon the importance of speech in early stages of design as it facilitates the ease of conveying ideas and overcome the disambiguation of the communication. Verbal communication also serves as a primary discourse for ideas in a collaborative setup [23, 24, 25]. Several workflows and interfaces have been designed and developed to leverage the communication within groups. For example, the *IdeaWall* presented by Shi et al. [26] supports group brainstorming by extracting essential verbal contents and providing real-time combinatorial visual cues. On similar lines, Chandrasegaran et al. [27] built on the notion of "smart meeting spaces" and proposed *TalkTraces* which captures and visualizes verbal contents in meetings. With the advancements in NLP, speech has become a more prominent way of interacting with automated systems. However, further exploration is needed to properly recognize users' intent in free-form conversation and implement suitable processing techniques such as context-aware topic modeling to enable an intuitive and effective communication environment [28].

2.2 Cognitive-Supported Workflows for 3D Ideation

Recently, there has been significant development in the direction of sketching-based 3D ideation. Often, the geometric modeling approach in these works is constrained such that the interface becomes a natural extension of how we think while sketching. As a result, most of these works are either grounded in tablet-based design interactions [29] akin to pen-and-paper sketching or direct 3D ideation through spatial inputs [30]. The key contribution for these works lies in cognitive-supported workflows primarily based on gestural actions for design ideation and exploration tasks. Prior works (§2.1) have highlighted the role of verbalization in design towards an uninhibited, intuitive and rapid design experience. To the best of our knowledge, a handful of works focus on design verbalization through speech-based interfaces for 3D modeling. In their work *MozArt*, Sharma et al. [14] showcase a multi-modal (touch and speech) interface for conceptual 3D-modeling where the primary role of speech is to command actions that are generally controlled through menu-based GUI inputs. Similarly, Plumed et al. [31] showcase a speech-based annotation workflow for computer-aided design on Solidworks. However, few recent works [16, 17] have begun to investigate the fundamental requirements for speech-based CAD interfaces with primary focus on designers from the domains of engineering and architecture. Most works on speech-based design interfaces are at a nascent stage of research exploration and therefore, it is imperative to conduct a systematic investigation on speech as a mode of design communication; the vision for it being used for 3D modeling akin to gestures and pen-based design inputs.

3 CONCEPTUAL OVERVIEW

Our goal for this work was to systematically investigate and operationalize speech-assisted workflows for 3D

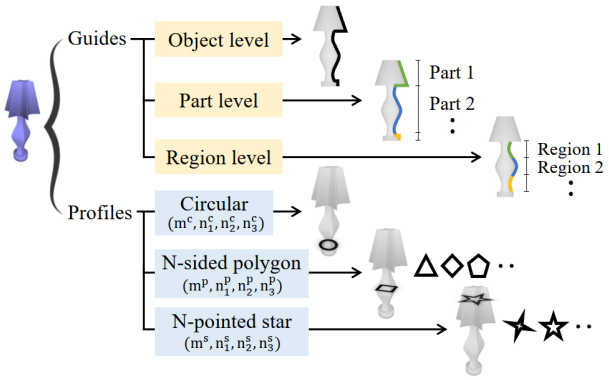

Fig. 2. Anatomical representation of our speech-based workflow for 3D shape exploration (illustrated using the lamp as a reference).

shape exploration through a specific case-based study of scene modeling. With this in mind, we designed a constrained 3D scene modeling application that allows users to explore 3D shapes through verbal inputs in the context of a scene comprising of seed shape templates. We discuss the key concepts involved in the design of our system below.

3.1 Shape Representation

One of the main challenges in our work was to determine an appropriate shape representation that could be amenable to (1) semantic mapping at multiple levels of detail (object, part, region) and (2) generalization beyond the specific system and interface decisions taken in this work. Based on these requirements, we propose to use lofts as our shape representation. In this work, we assume that each **object** (e.g. a lamp) is represented as a single lofted volume that can be semantically segmented into multiple **parts** (e.g. lamp-shade, stem, base) which can further be divided into semantic **regions** of interest as appropriate (e.g. the stem of the lamp may be modified independently at the top, middle, or bottom regions). Lofts offer an elegant way to decompose the description of a shape [30] in terms of two curves (Figure 2): section and guide curves, that could be mapped easily to simple verbal descriptions making them ideal for modeling a wide range of real-world objects.

3.2 Our Chosen Case Study: Interior Scene Design

Our speech-assisted workflow is designed on the basis of the four areas of focus: (1) design-in-context, (2) order independence, (3) multi-scale exploration, and (4) semantic regions of interest. The idea is to enable users to explore shapes for multiple objects in a given scene, parts constituting those objects, and regions within parts. In our work, we specifically chose an interior design context composed of a scene containing three pots and a lamp on a table.

Design-in-Context: Providing a context to our users is crucial in aiding their exploration process as well understanding their design choices. Our scene, therefore, enables users to form contextual relations between individual objects. This is an important form of interaction in our workflow, since users design choices can change based on other objects. Our workflow makes these changes observable by dynamically changing the locations of the objects on the table based on the size of the table. For instance, making the table smaller automatically brings the objects closer.

Order Independence: Our workflow allows users to freely explore any object, part and region independent of order. Users can simply choose the toggle corresponding to their desired object and make modifications. They can also undo their actions and reset to default, ensuring no constraints on a prescribed sequence.

Multi-scale Exploration: We allow users to create primitive shapes (e.g., pentagon), and then add finer details to it, such as making it *'starry'* (5-pointed star). Additionally, users have the ability to directly make finer detailed shapes by simply describing them that way. We follow a similar approach for changing dimensions, where generic terms such as *'bigger'* or *'smaller'* can be used to change dimensions, and more precise/relative changes can be made by describing the percentage of change. This ability to create hierarchical and direct modifications gives users more freedom to naturally describe shapes.

Semantic Regions-of-Interest: Shapes such as pots are typically single-body objects that contain several regions, such as a neck or a mouth, that when modified, can create a variety of different looking pots. To make these changes possible through speech, we semantically divide and label the objects into different regions. Pots contain a: mouth, neck, body and base. Lamps and tables, being multi-component objects are first divided into parts, and then into regions (for specific parts only). This allows users to create a wider variety of designs for the same objects.

4 SYSTEM IMPLEMENTATION

4.1 Software & Hardware Setup

We developed our user interface using Unity3D Game Engine scripted with C# language. For our speech-to-text recognition and intent classification modules, we used Azure's Custom Speech API and Language Understanding (LUIS) API. Our interface was deployed on a laptop with i7-4720HQ processor, 8GB RAM, and a NVIDIA GTX 950M GPU. Built-in microphones were used to record speech.

4.2 Scene Modeling

We utilize a loft-based 3D modeling technique to render our shapes. Lofts allow us to explore a wide variety of shapes by changing their 2D sections and guide curves at different locations.

4.2.1 Section Geometry

We use the superformula equation to generate 2D profile shapes for the lofts' sections. The superformula equation is a generalization of a superellipse and can generate a wide variety of shapes by changing its parameter values. The equation is given below:

$$r(\phi) = \left(\left| \frac{\cos(\frac{m\phi}{4})}{a} \right|^{n_2} + \left| \frac{\sin(\frac{m\phi}{4})}{b} \right|^{n_3} \right)^{\frac{1}{n_1}}$$
$$x = r(\phi) \times \cos(\phi)$$
$$y = r(\phi) \times \sin(\phi)$$

Here, r and ϕ are the radius and angle of the superformula shape in polar coordinates. m defines the number of corners of the shape, while n_2 and n_3 determine if the shape is inscribed or circumscribed within the unit circle ($a = b = 1$). n_1 can change sharpness of corners and curvature of sides. By keeping the values of n_2 and n_3 equal to each other and below 2, symmetric shapes can be generated. The superformula, therefore, provides a convenient way to generate a wide variety of shapes by changing parameter values (Figure 4).

4.2.2 Guide Curves

Guide curves help us define the basic outline of the 3D shapes of the objects (Figure 3). We utilized a combination of linear and cubic spline curves to model our guide curves. However, in order to maintain the semantic identity of each object in our scene, we added geometric constraints to these splines. For instance, while the guide curve for the pot (a single part object) is a single cubic spline, the guide curve for the lamp and table is a combination of straight lines and cubic splines. Furthermore, we also introduce "padding points" along the splines to be able to define semantic regions (e.g. neck, body on the pot) on these objects.

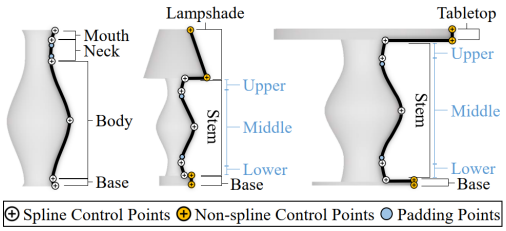

Fig. 3. Guide curves for the pot, lamp and table are shown. Guide curve of pot, and stems of the lamp and table are defined by spline curves.

4.2.3 Section Smoothing

We utilize Laplacian smoothing to enable users to generate smooth cross-sections. Given a region defined by the user on the object, we transform each section in the region on to a plane and apply smoothing in the plane before transforming the section back to 3D space. The smoothing algorithm is as follows: p_i = 0.5 x (p_{i-1} + p_{i+1}). Here, p_i is the smoothed vertex for neighboring vertices p_{i-1} and p_{i+1}.

4.2.4 Recalculating Object Location in Context

In our scene, all objects are located on a table. Therefore, we implemented an automated method to re-position each object whenever the table is edited. This is an important aspect for studying design-in-context.

4.3 Semantic Mapping of Shapes & Dimensions

A key challenge we faced was semantically mapping the user's speech to different shapes and sizes. The superformula afforded us a convenient way to map shape descriptions of primitive polygonal shapes (e.g., triangle, square, etc.) to their respective shape parameters (Figure 4). We then allowed users to add finer details to these shapes through commands such as "*starry*" (by reducing n_1) and "*smooth*" (using Laplacian smoothing). In this work, we limited the number of polygonal vertices to a hexagon; mainly to simplify the shape exploration user experience. However, expanding the shape vocabulary to more shapes is a fairly straightforward task.

Additionally, users could also make two types of dimensional changes to the objects: *coarse modifications* and *relative size modifications*. Coarse modifications

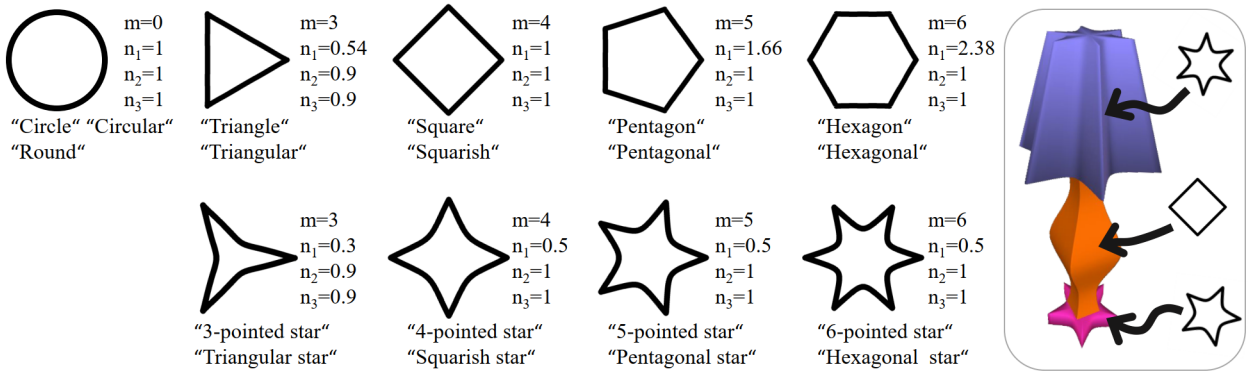

Fig. 4. Semantic mapping of shapes to their superformula parameters. The different labels and parameter values (m, n_1, n_2 and n_3) for each shape are provided. An example of the cross-sectional superformula shapes of a lamp are shown (Right).

were used to make general dimensional changes, such as making the region "*bigger*", or "*shorter*". Relative size modifications were used to describe percentages by which coarse modifications should be made, for instance, making a part "*30% smaller*". These changes were mapped to the position of the control points of the guide curves.

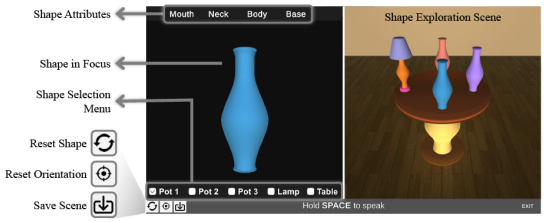

Fig. 5. User Interface (UI) of our speech-based system is shown. Users can select objects they want to explore by choosing their respective toggles. They can simultaneously view changes in the scene to right. Interaction with system happens by holding down the space bar and speaking commands which are transcribed and shown in the bottom gray panel.

4.4 Speech Recognition and Intent Classification

We implemented our speech-to-intent classification task in two steps: (1) transcribing the user's spoken commands and (2) classifying the transcribed commands into specific actions. For the first task, we used Microsoft Azure's Cognitive Speech Service to train a custom speech model catering to our specific application. We trained the model using audio and text data from 6 different users, containing utterances commonly used for our application.

To classify transcribed texts into scene modeling actions, we used Microsoft Azure's Language Understanding (LUIS) service to train custom models that could predict the overall meaning of the users' commands. The model required two components: *Intents* (specific actions) and *Entities* (features corresponding to the actions). We defined two main intents: changing sectional shapes and changing dimensions, and both of these shared one common entity, i.e., the part/region to be changed. The part entity consisted of a list of all the parts and regions present in our scene. The entity specific to the shape changing intent consisted a list of all the shapes that our system could represent, and similarly, the entity specific to the dimension changing intent contained all types of dimension changes that were allowed by our system. Next, we trained the model using 35 example utterances. As the LUIS models are built on pre-trained NLP models, the number of example utterances necessary to get high prediction accuracy was quite low, specially with the added feature lists. The accuracy of our trained model was consistently above 96% for test utterances.

4.5 Interface Elements

We designed a split-screen visualization for our interface (Figure 5). This approach allows users to focus on the objects they want to modify (detail) while also being able to see their changes reflect in the scene (overview).

To interact with the system, users need to press and hold the space bar while speaking. Rotation and zooming into the object is possible using a mouse and we also provide buttons for refreshing, resetting views and saving scenes, reducing the commands that users need to remember. Another visual feature we added was to highlight regions/parts that the users intended to change by changing their color for a second, after receiving the user's commands. This helped users clearly identify their changes.

5 EXPERIMENT DESIGN

We designed a case study of interior design, where users can explore scenes by making changes to individual objects within the scene using their speech. The scene-creation task would help us understand the key concept of design-in-context for speech-based design workflows.

5.1 Participants

We recruited 6 participants (4 male, 2 female) belonging to the age group of 18-30 years. Participants were enrolled in undergraduate and graduate degrees with backgrounds in engineering, architecture, and visualization. Five participants had prior design experience through research and coursework and were familiar with 3D modeling tools such as AutoCAD, Solidworks, Maya, etc.

5.2 Procedure

Each study lasted between 60-70 minutes. Participants were provided with a pre-study questionnaire eliciting their experiences with speech-based systems and CAD tools. Next, they were introduced to our design interface and were asked to perform a practice task (15-20 minutes) with a scene comprised of one pot, lamp and a table.

The main study task (30 - 40 minutes) was to iterate and design a variety of scenes starting from the default template. Participants were instructed to explore different scenes by modifying the individual objects. We did not limit the maximum number of scenes they could create, but encouraged them to explore at least three different scenes. Participants were asked to save their scene whenever they were satisfied with their changes. Finally, participants completed a post-study questionnaire containing the creativity support index [32], system usability scale [33] and general feedback. For each study, we recorded the transcribed speech commands; user intents; 3D scene mesh; and video recording of the screen.

6 FINDINGS

Participants reported an overall positive user experience and were able to make a wide variety of design changes to objects in the scene. Each participant explored an average of 138 variations (max: 161, min: 118) per study session. This included changes to the object shape (avg.: 49, max: 77, min: 33) and physical dimension (avg: 89, max: 110, min: 77). We discuss our observations in detail in the following sections.

6.1 Design-in-Context

In general, we observed participants making changes to the objects "in context" to the scene. For instance, one participant explained that they were trying to create a *soda bottle* and a *candle stick* as they imagined these items to be on a *dining table* (User 6 in Figure 6). They also increased the size of the table to make the scene "less cluttered". When asked about the effect of other objects in the scene, they answered: *"Yes, I would have liked to place the lamp in the middle of the table if the other objects were not present"*. Another participant expressed their desire to explore unique varieties of shapes across all objects and described their pot with a thin and long neck as one that could be used to display flowers (User 4 in Figure 6). We received similar feedback from other users highlighting the importance of context to the design of objects in the virtual scene.

6.2 Order Independence

While no general trend was observed across all users with respect to modification sequences; we did observe certain user-specific patterns which we explain in two parts:

6.2.1 Object Level Sequence

Three participants edited each of the five objects at least once before modifying previously explored objects while two others did the same for four objects. This might suggest that participants preferred modifying newer objects before returning to already explored objects. Another observation regarding the sequence was that half the participants started their scene creation task by modifying the table first, suggesting that they preferred setting the reference base for the objects placed on top of it. The other half modified Pot 1 first, which was the default starting point for the scene creation task. Additionally, we observed majority of participants modifying the pots sequentially in order of their label numbers, possibly due to the orientation of the pots in the scene or the order of toggle buttons.

6.2.2 Parts & Regions

In regards to parts/regions of objects, users preferred design modifications to create uniformity between similar objects (i.e. Pots). Two users modified the bodies of all their pots before modifying other regions, while another modified the bases first. One user also followed a similar design modification sequence for two pots, making them visually similar (third row Figure 6). Users also preferred modifying regions of a single part consecutively

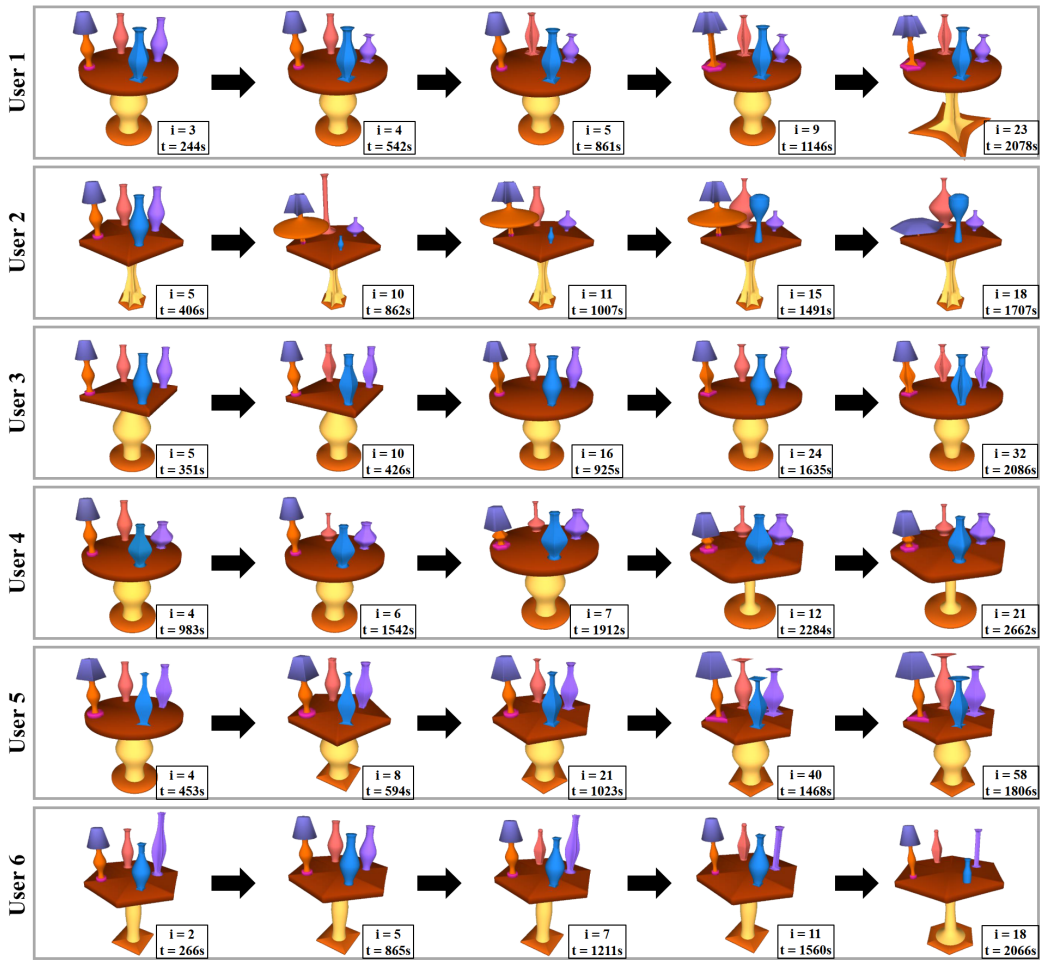

Fig. 6. Scenes created by all six users at different time frames of their study (each row corresponds to one user). These scenes were specifically selected to show variety of shapes explored by each user. Iteration number(i) and timestamp(t) shown on the bottom right of each scene.

before moving to other parts. For instance, five participants modified multiple regions of the stem of tables and lamps before moving to other parts.

6.3 Multi-scale Exploration

Most users took advantage of the multi-scale design modifications afforded by our interface. Four participants preferred only the hierarchical approach of creating a primitive shape followed by detailed exploration, while two others preferred making direct detailed modifications. We observed a flipped approach for dimensional modifications, where four participants preferred making fine changes (relative percent changes) more than 60% of the time. The preference for relative size changes must have been due to greater degree of visual changes when compared to coarse changes. Making parts/regions bigger was the most frequent user request for dimension changes, while making them taller was the least requested.

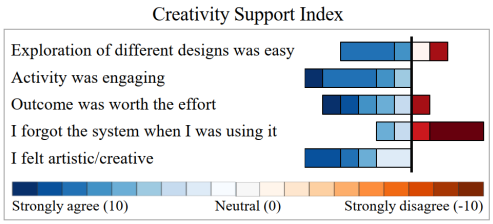

Fig. 7. Creativity Support Index

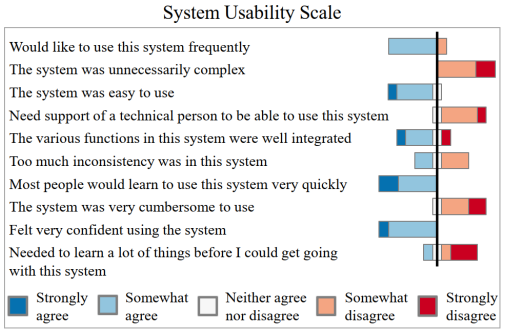

Fig. 8. System Usability Scale

6.4 Semantic Regions of Interest

In our post-study questionnaire, most users mentioned that the pre-defined regions were intuitively labeled. This was evident in their design changes, where all users preferred modifying specific regions of the pots instead of the entire pot, with five participants making close to a 100% of their modifications on just regions. We saw a similar trend with the stems of the lamp and table, where four users made an excess of 60% of their design modifications to the lamp's stem regions and five users made majority of their modifications to the table's stem regions, when compared to the entire stem. These observations help us understand the role of semantic regions in aiding design exploration.

6.5 User Feedback

Participants found our system fairly easy to use and felt confident doing so (Figure 8). Participants also found the activity engaging and felt creative while being able to explore different designs easily (Figure 7). One participant with an architecture background shared that a system like ours could help architects and clients come up with design solutions quickly, without delayed reviews.

6.6 Limitations

An important limitation of our interface was the limited vocabulary. Although our speech model could understand natural conversations, it required training on shape-specific terms. While limiting, this constraint was intentional to minimize confusion for the users making specific modifications and helped them stay focused on the exploration task. This could be a reason why users felt our system was inconsistent. We can mitigate this in future interfaces by conducting design elicitation studies similar to recent works [16, 17] and using it to train our speech model. Some participants also felt limited by the choice of shapes, which we purposely intended, to avoid overloading users with extra terms. We could easily expand on this once we integrate a more natural language approach.

7 DISCUSSION & FUTURE DIRECTIONS

7.1 Hierarchical and Detailed Changes

Most users preferred making hierarchical (overview-then-detail) changes to shapes and precise changes to dimensions. For future speech-based workflows, we can explore more ways of offering hierarchical changes to the shapes (e.g., sharpen, chamfer), and relative or unit-based changes for dimensions (e.g. half the size, inches).

7.2 Multi-Context Approach

The interior design context was a driving factor for most users' design modifications. Providing other forms of context, such as configuration of objects (relative locations, scales) and shape forms can improve the user interaction with our workflow. For instance, users could ask to make the shape of one object same as another object. This reference-based exploration can be advantageous in speech-based workflows due to the ease of contextual descriptions in spoken language.

7.3 Free-Form Region Exploration

In our work identifying and semantically dividing objects into intuitive regions of interest was a challenging task which paid off, as most users preferred making changes to these regions. However, allowing users to define and label their own regions could make the design process more personalized while also allowing users to modify arbitrary shapes.

7.4 Future Directions

What we show in this paper is merely a glimpse of what could be possible in speech-based shape modeling. We see tremendous opportunities for integrating direct manipulation, sketching, and user-provided region selection with speech-based approaches. Collaborative interfaces would also be an obvious area of extension and exploration beyond our work. However, what is most intriguing to us is the possibility for developing a scalable and general shape representation that could allow the integration of a whole host of tools in the linguistic toolbox such as metaphors, similes, and analogies into digital tools for 3D shape ideation.

REFERENCES

[1] Schon, D. A., 1984, *The reflective practitioner: How professionals think in action*, Vol. 5126 Basic books.

[2] Barelkowski, R., 2010, "Verbal thinking in the design process: Internal and external communication of architectural creation.," *Design Principles & Practice: An International Journal,* **4**(5).

[3] Ranscombe, C., Bissett-Johnson, K., Mathias, D., Eisenbart, B., and Hicks, B., 2020, "Designing with lego: Exploring low fidelity visualization as a trigger for student behavior change toward idea fluency," *International Journal of Technology and Design Education,* **30**, 04.

[4] Jonson, B., 2005, "Design ideation: the conceptual sketch in the digital age," *Design Studies,* **26**(6), pp. 613–624.

[5] Sosa, R., 2019, "Accretion theory of ideation: evaluation regimes for ideation stages," *Design Science,* **5**, 11.

[6] Ungureanu, L.-C., and Hartmann, T., 2021, "Analysing frequent natural language expressions from design conversations," *Design Studies,* **72**, p. 100987.

[7] Dossick, C. S., and Neff, G., 2011, "Messy talk and clean technology: communication, problem-solving and collaboration using building information modelling," *The Engineering Project Organization Journal,* **1**(2), pp. 83–93.

[8] Bolt, R. A., 1980, ""put-that-there" voice and gesture at the graphics interface," In Proceedings of the 7th annual conference on Computer graphics and interactive techniques, pp. 262–270.

[9] Chu, C.-C. P., Dani, T. H., and Gadh, R., 1997, "Multi-sensory user interface for a virtual-reality-based computeraided design system," *Computer-Aided Design,* **29**(10), pp. 709–725.

[10] Gao, S., Wan, H., and Peng, Q., 2000, "An approach to solid modeling in a semi-immersive virtual environment," *Computers & Graphics,* **24**(2), pp. 191–202.

[11] Weyrich, M., and Drews, P., 1999, "An interactive environment for virtual manufacturing: the virtual workbench," *Computers in industry,* **38**(1), pp. 5–15.

[12] Nanjundaswamy, V., Kulkarni, A., Chen, Z., Jaiswal, P., Verma, A., and Rai, R., 2013, "Intuitive 3d computer-aided design (cad) system with multimodal interfaces," In International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Vol. 55850, American Society of Mechanical Engineers, p. V02AT02A037.

[13] Kou, X., Xue, S., and Tan, S., 2010, "Knowledge-guided inference for voice-enabled cad," *Computer-Aided Design,* **42**(6), pp. 545–557.

[14] Sharma, A., Madhvanath, S., Shekhawat, A., and Billinghurst, M., 2011, "Mozart: a multimodal interface for conceptual 3d modeling," In Proceedings of the 13th international conference on multimodal interfaces, pp. 307–310.

[15] Wiegers, T., Langeveld, L., and Vergeest, J., 2011, "Shape language: How people describe shapes and shape operations," *Design studies,* **32**(4), pp. 333–347.

[16] Khan, S., and Tunçer, B., 2019, "Speech analysis for conceptual cad modeling using multi-modal interfaces: An investigation into architects' and engineers' speech preferences," *Artificial Intelligence for Engineering Design, Analysis and Manufacturing : AI EDAM,* **33**(3), 08, pp. 275–288.

[17] Khan, S., Tuncer, B., Subramanian, R., and Blessing, L., 2019, "3d cad modeling using gestures and speech: Investigating cad legacy and non-legacy procedures," In Proceedings of the 18th International Conference, CAAD Futures 2019, CUMINCAD.

[18] Khan, S., and Tunçer, B., 2019, "Gesture and speech elicitation for 3d cad modeling in conceptual design," *Automation in Construction,* **106**, p. 102847.

[19] Wetzstein, A., and Hacker, W., 2004, "Reflective verbalization improves solutions—the effects of question-based reflection in design problem solving," *Applied Cognitive Psychology,* **18**(2), pp. 145–156.

[20] Hong, Y.-C., and Choi, I., 2011, "Three dimensions of reflective thinking in solving design problems: A conceptual model," *Educational Technology Research and Development,* **59**(5), pp. 687–710.

[21] Adler, A., and Davis, R., 2007, "Speech and sketching for multimodal design," In *ACM SIGGRAPH 2007 Courses*, SIGGRAPH '07. Association for Computing Machinery, New York, NY, USA, p. 14–es.

[22] Adler, A., and Davis, R., 2007, "Speech and sketching: An empirical study of multimodal interaction," In Proceedings of the 4th Eurographics workshop on Sketch-based interfaces and modeling, pp. 83–90.

[23] Allen, K., 2005, "Online learning: constructivism and conversation as an approach to learning," *Innovations in Education and Teaching International,* **42**(3), pp. 247–256.

[24] Nijstad, B. A., and Stroebe, W., 2006, "How the group affects the mind: A cognitive model of idea

generation in groups," *Personality and social psychology review,* **10**(3), pp. 186–213.

[25] Wang, H.-C., Cosley, D., and Fussell, S. R., 2010, "Idea expander: Supporting group brainstorming with conversationally triggered visual thinking stimuli," In Proceedings of the 2010 ACM conference on Computer supported cooperative work, pp. 103–106.

[26] Shi, Y., Wang, Y., Qi, Y., Chen, J., Xu, X., and Ma, K.-L., 2017, "Ideawall: improving creative collaboration through combinatorial visual stimuli," In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pp. 594–603.

[27] Chandrasegaran, S., Bryan, C., Shidara, H., Chuang, T.-Y., and Ma, K.-L., 2019, "Talktraces: Real-time capture and visualization of verbal content in meetings," In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14.

[28] Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J., et al., 2019, "The state of speech in hci: Trends, themes and challenges," *Interacting with Computers,* **31**(4), pp. 349–371.

[29] Piya, C., , V., Chandrasegaran, S., Elmqvist, N., and Ramani, K., 2017, "Co-3deator: A team-first collaborative 3d design ideation tool," In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17, Association for Computing Machinery, p. 6581–6592.

[30] Vinayak, Ramanujan, D., Piya, C., and Ramani, K., 2016, "Mobisweep: Exploring spatial design ideation using a smartphone as a hand-held reference plane," In Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, TEI '16, Association for Computing Machinery, p. 12–20.

[31] Plumed, R., González-Lluch, C., López, D., Contero, M., and D. Camba, J., 2020, "A voice-based annotation system for collaborative computer-aided design," *Journal of Computational Design and Engineering*, 12.

[32] Carroll, E. A., and Latulipe, C., 2009, "The creativity support index," In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09. Association for Computing Machinery, New York, NY, USA, p. 4009–4014.

[33] Brooke, J., 1996, "Sus: a "quick and dirty'usability," *Usability evaluation in industry,* **189**.