Contributors to the aesthetic judgement of 3D virtual sculptures

Edward Easton
Department of Computer Science,
Aston University, Birmingham, UK
eastonew@aston.ac.uk

Ulysses Bernardet
Department of Computer Science,
Aston University, Birmingham, UK
u.bernardet@aston.ac.uk

Anikó Ekárt
Department of Computer Science,
Aston University, Birmingham, UK
a.ekart@aston.ac.uk

ABSTRACT

Aesthetic judgement plays a key role in many aspects of everyday life, judging an object to be aesthetically pleasing often heightens the pleasure and enjoyment derived from that object. One area where this applies is artwork, where most genres and styles of art heavily rely on being considered aesthetically pleasing. It has been shown that an aesthetic judgement of a piece of art combines many different aspects, all contributing to the assessment. Identifying and understanding these aspects for 2D images has been extensively investigated, however, 3D items have not been considered to the same degree. In this paper, we investigate which aspects contribute to the aesthetic judgement of 3D virtual sculptures, using a gamified approach within a custom VR environment. Participants were able to express which aspects contributed to their assessment of the virtual sculptures. We found that some stalwart 2D aspects, such as complexity and order, are not as highly important for 3D items, being replaced by other characteristics such as how dynamic the sculpture appeared.

CCS CONCEPTS

• **Applied computing** → Arts and humanities; Media arts; • **Human-centered computing** → Human computer interaction (HCI); Interaction paradigms; Virtual reality; Human computer interaction (HCI); Empirical studies in HCI; • **Computing methodologies** → Artificial intelligence.

KEYWORDS

Aesthetic judgement, 3D Art Generation, Gamification, Virtual Reality

ACM Reference Format:
Edward Easton, Ulysses Bernardet, and Anikó Ekárt. 2023. Contributors to the aesthetic judgement of 3D virtual sculptures. In *11th International Conference on Digital and Interactive Arts (ARTECH 2023), November 28–30, 2023, Faro, Portugal*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3632776.3632787

1 INTRODUCTION

Aesthetic judgement is a multi-faceted process with innumerable aspects contributing to the assessment [18], [22], [35], [41]. The judgement itself relates to whether an individual likes a certain item, it transcends both traditional viewpoints where an aesthetically pleasing piece of art is also considered beautiful and other perspectives where the beauty of the item is entirely unrelated. On both sides, the judgement remains the same and concludes with whether the item being judged is liked or not, with only the aspects contributing to the judgement differing. Some aspects, that contribute to aesthetic judgement are relatively well-known such as the order and complexity [7], [13], [20], however, many of the aspects are left in relative obscurity. One common problem with identifying these aspects is that it can be difficult for someone to articulate why they find a piece of art aesthetically pleasing. Whilst this process becomes easier with practice, a higher level of expertise in art can influence how items are judged [23], [30]. Experts can provide a wealth of information, however, without gaining further knowledge on which aspects contribute to aesthetic judgement the aesthetic judgement process is difficult to fully model.

Whilst non-experts may find describing their judgement more difficult, aesthetic judgement still occurs, they will still like or dislike an object. Currently, experts are often used to help model the aesthetic process, however, to truly understand the process, nonexperts and everyone in-between need to be included in the model, paving the way for computer systems which can generate artwork that appeals to a wider variety of people, expanding the reach, impact and potential uses these systems can have.

One way to approach modelling aesthetic judgement is to start by determining which terms are commonly used to describe artwork, rather than starting with the creation of the artwork itself. These terms are the natural way to describe artwork and aesthetic concepts and provide an entry point for people who are less familiar with describing artwork to describe what contributes to their aesthetic preferences. Finding which terms are used by non-experts provides a good starting point for understanding the aesthetic judgement process, eventually enabling the creation of artwork applicable to a wider range of people.

As more aspects contributing to aesthetic judgement have been identified, it will be possible to map which attribute or combination of attributes, when applied to a piece of art, will contribute to the artwork exemplifying the selected term.

Some of the difficulty of formalising the aspects lies with the differences between the computational understanding of these terms compared to how they are understood by humans. Computational approaches generally look at attributes which are easily understood by computers, often visual aspects that are easy for a computer to process, such as symmetry [2] and a variety of statistical analyses such as contrast [25] and naturalness [1]. This is not to say more complex aspects have not been investigated, for example, aspects like ethical considerations [9], [41], but the computational interpretation will not necessarily reflect the human understanding of the same term and generally does not allow for ambiguity in terms. As an example, the term complexity could relate to a piece of artwork having complex emotional content or complex subject

ARTECH 2023, November 28–30, 2023, Faro, Portugal

ACM ISBN 979-8-4007-0872-5/23/11
https://doi.org/10.1145/3632776.3632787

matter as well as visual complexity such as multiple colours, lines and shapes [26]. This indicates why a disparity exists between the human judgement and the analogue followed by auto-generation systems. Attempting to solve this disparity forms an important part of the auto-generation of aesthetically pleasing artwork and while this disconnect remains, it will be difficult for any generation system to reliably create aesthetically pleasing artwork.

Other approaches to the subject of aesthetic judgement concern themselves with the human understanding of the process and some cross-over between these and computational approaches do exist [17], but are under-utilised, requiring further investigation to merge the two different understandings of aesthetic judgement. We present an initial step towards merging the two approaches.

In the remainder of this paper, we first look at how the process can be analysed to extract the details on the aspects that contribute to aesthetic judgement. Next, we detail how the data was obtained, how the tags were chosen and how the environment was created to investigate the aesthetic concepts. Then we present the data that was collected, discuss the results, and draw conclusions and explain how we will use these results moving forward.

2 OBTAINING THE RELATIVE IMPORTANCE OF AESTHETIC ATTRIBUTES

The high variety of artwork reflects the complicated nature of aesthetic judgement. Out of all potential aspects, not all contribute to the same extent or even in the same way across different judgements in different contexts, suggesting that aesthetic judgement is not a static process [39]. The first step to addressing this is the attempt to identify which aspects form an important part of the aesthetic judgement process while accounting for the subjectivity in the terms and their application. Within the context of the auto-generation of art, it is important to understand which aspects contribute significantly to the aesthetic judgement process.

The Identified aesthetic terms would form the basis of generating artwork, however, to use them effectively more details are required such as the relative contributions of the terms, which terms provide the most useful information about the items they are describing and finally whether these terms contribute positively or negatively to the judgement. To collect this, item analysis was chosen as an effective approach that is often applied to designing exam questions [6], [36], [40]. With a few minor amendments item analysis can also be applied to help understand the different aspects which contribute to the aesthetic judgment of artwork.

The tags that are the most commonly applied when judging artwork will represent the popularity of the terms being applied.

The more popular the tag, the more important it will be to aesthetic judgement. As well as the popularity, the consistency of the application of the tags is also considered. This protects against choosing ambiguous terms and ensures that enough data will be collected to help formalise the aspect.

To calculate both consistency and popularity, the allocation of the tags to sculptures by participants needs to be defined, as shown in Equation 1:

$$allocation_{p\ t\ s} = \begin{cases} 1\ participant\ p\ assigns\ tag\ t \\ \quad to\ sculpture\ s \\ 0\ otherwise \end{cases} \tag{1}$$

Using the defined allocations, the mean assignment of each tag t to a sculpture A_t can be calculated, by dividing the total number of times the tag has been assigned to each sculpture by the number of sculptures as shown in Equation 2:

$$A_t = \sum_{s=1}^{S} \sum_{p=1}^{P} \frac{allocation_{p\ t\ s}}{S} \tag{2}$$

The consistency value for each tag and sculpture can then be calculated by dividing the mean assignment for a tag (A_t) by the total number of participants (P), as shown in Equation 3. The consistency indicates whether a specific tag has been reliably applied to sculptures which exhibit the same attributes and that the link between these attributes and the applied term is identifiable by the participants.

$$C_t = \frac{A_t}{P} \tag{3}$$

Two further details can also be obtained about each tag: how difficult the tag was to apply and how discriminating the tag is. A highly discriminating tag will be chosen less frequently against sculptures which do not represent the tag and tags with a high difficulty will not be applied frequently. Both of these measures reveal important information: the difficulty of the application allows the discarding of tags which are too difficult or too easy to apply as they provide too little or too much information on which pieces of art exhibit the term. The discrimination allows the identification of tags which are representative of the sculptures they are applied to, items which do not discriminate enough would be an indication that the tag is ambiguous and can be mapped to multiple attributes. We consider the tags which are identified as having a medium level of application difficulty and a high level of discrimination would be suitable candidates for further analysis and implementation in art-generating systems. By setting these thresholds for both the difficulty and discrimination values obtained for each tag, a shortlist of aspects can be obtained where each item has sufficient data collected within this experiment in order to potentially learn what physical attributes map to the abstract aesthetic terms.

$$E_t = A_t \tag{4}$$

The endorsement E_t of a tag (Equation 4) is the mean number of times it has been applied to each item. It can be interpreted as how easily a tag can be applied to the piece of art, where the higher the number of average applications, the easier the tag is to apply. We consider the difficulty of application to be the complement of ease of application and as such we calculate it as the involution function.

$$Diff_t = 1/A_t \tag{5}$$

For the discrimination to be evaluated, first a total score must be calculated for each piece of art, set as the sum of applications of all tags applied by all participants to the artwork:

$$Score_s = \sum_{t=1}^{T} \sum_{p=1}^{P} allocation_{p\ t\ s} \tag{6}$$

Calculating the discrimination and whether the term has a positive impact on aesthetic judgement requires calculating the correlation between the sculptures and the calculated data. The correlation[1] (Equation 9) is calculated using the standard deviation[2] (σ), shown in Equation 7, and the covariance[3] (s_{xy}), shown in Equation 8, representing how the two sets of data are linearly related.

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{N}} \tag{7}$$

$$s_{x\ y} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \hat{x})\,(y_i - \hat{y}) \tag{8}$$

$$r_{x\ y} = \frac{s_{x\ y}}{\sigma_x\ \sigma_y} \tag{9}$$

The discrimination value for each tag is calculated as the correlation between the total number of times the tag has been assigned to the artwork and the total number of tags assigned to the artwork, if a sculpture represents a particular tag it will have a higher correlation between these values.

$$D_t = r_{x\ y}\left(\sum_{p=1}^{P} allocation_{p\ t\ s},\ Score_s\right) \tag{10}$$

The final measure, which provides useful information about the terms considers the positive or negative influence that each tag has on the aesthetic judgement of an artwork, requires the overall aesthetic rating for each item (Equation 11). Each participant was asked to rate the sculpture between 0 and 10 (*ratings*), this is correlated to the total applications of each tag on that sculpture, which indicates whether the tag is considered as a positive or a negative aspect of the judgement, the more times a tag has been assigned to a sculpture with a higher rating, the higher the positive impact that that tag has.

$$Pos_{t\ s} = r_{x\ y}\left(rating_s,\ \sum_{p=1}^{P} allocation_{p\ t\ s}\right) \tag{11}$$

3 METHODOLOGY

It can be easy to determine whether someone likes a sculpture by collecting ratings for each item, an approach which has been used extensively to judge auto-generated artwork to test new generation methods or to investigate whether a particular feature contributes to aesthetic judgement. To collect the required data to determine contributing factors, several aspects need to be considered: what terms the user can use to describe the artwork, what artwork to display to a user and how the user will be able to describe the artwork.

3.1 Concept Tags

Determining which terms to include is made more difficult due to several restrictions which need to be applied: (1) terms should not be too technical to ensure they can be understood by all participants. (2) As much of the conceptual aesthetic space as possible needs to be covered to try and fully explore how individuals judge sculptures.

[1]https://www.r-tutor.com/elementary-statistics/numerical-measures/correlationcoefficient
[2]https://www.r-tutor.com/elementary-statistics/numerical-measures/standarddeviation
[3]https://www.r-tutor.com/elementary-statistics/numerical-measures/covariance

Table 1: Aesthetic terms

Term	Value	Term	Value
Quiet	0.17	Ordered	0.18
Boring	0.19	Complex	0.19
Dynamic	0.21	Simple	0.23
Stiff	0.25	Balanced	0.25
Disconnected	0.25	Unified	0.25
Irregular	0.27	Natural	0.27
Plain	0.28	Gentle	0.28
Neutral	0.29	Exciting	0.29
Interesting	0.29	Dull	0.32
Surprising	0.32	Alive	0.38
Freakish	0.38	Graceful	0.38
Lifeless	0.38	Messy	0.38
Practical	0.38	Predictable	0.38
Strange	0.38	Strong	0.38
Controlled	0.38	Ugly	0.38
Unemotional	0.38	Unfriendly	0.38
Unnatural	0.38	Unpleasant	0.38
Weak	0.38		

To compile the initial list of terms, an overview of aesthetic judgement was sought. Work undertaken by Sibley [37], who investigated which terms were commonly used to describe artwork by art critics, founded the basis of our search, resulting in a list of 134 terms. Due to their origin, these terms are inherently applicable to the judgement of artwork, and whilst they were compiled a long time ago, their meanings are still relevant in this context. To avoid overwhelming the participants with too many terms, the list was filtered further to include only 30 terms, by checking the semantic similarity using the WordNet2[4] database. This allowed the reduction of the list of terms whilst still maintaining the coverage of a wide range of the conceptual space. The similarity was calculated by measuring the distance to the lowest common linked word between each pair of terms, items which have a closer common ancestor are considered more similar. Once the value had been calculated between each pair of terms, an average was taken for each one to represent how generally different that term was from the others, examples of the terms and their average semantic similarity are shown in Table 1.

The terms were then manually sanitised to ensure the list contained no antonyms, synonyms or duplicates and ensure that the term did not use outdated language, where the meaning of the term may not be obvious to all participants e.g. Gaudy. Finally, any terms which were directly connected to emotional state were removed, for example, happy or sad, as their inclusion may unintentionally affect the results due to differences in participants' emotional state before taking part in the experiment. The top 14 dissimilar terms were selected, then to widen the area of semantic space being covered by this reduced list, their antonyms were added to the list to obtain the final list of aesthetic concepts, shown in Table 2. These terms represent the widest area of the aesthetic concept space available using

[4]https://wordnet.princeton.edu/

Table 2: Final list of aesthetic terms

Final list of aesthetic terms		
Busy	Cold	Quiet
Angular	Unrefined	Ordered
Complex	Boring	Drab
Friendly	Static	Separate
Calm	Original	Natural
Curved	Warm	Loud
Simple	Sophisticated	Disordered
Unfriendly	Interesting	Bright
Unoriginal	Dynamic	Connected
	Unnatural	

the chosen inputs whilst removing as much subjectivity as possible. It should be noted that while these terms were removed here, they may still be suitable for investigation using a more suitable experimental paradigm.

3.2 Sculptures

In addition to the tags needing to cover a wide area of the aesthetic concept space, the same criterion needed to be applied to the sculptures that were displayed to the user.

Multiple ways exist to auto-generate 3D items including manipulating the colour and visibility of voxels using CPPNs [19] or Context-Free Grammars [4], [29], [34], using Graph Grammars to generate a set of points in 3D space [28], using Shape Grammars to create 3D items [10], [31], [33] and evolving parameters to use within an external generation system [27], [32]. However, to ensure that a wide range of visually different sculptures was created, the Axial Generation Process (AGP) was selected [16], which places geometric items around a central axis and is capable of creating a wide range of visually different sculptures. This process was also chosen as it generates sculptures which are suitable for use within Virtual Reality. For this experiment, the sculptures were created by placing 475 spheres or cubes within the bounds of a 1x1x2m containing box.

To ensure that the sculptures were visually different from each other and display a wide variety of attributes, a distance search was implemented using a Genetic Algorithm, with a fitness function which determined how dissimilar one sculpture was from another. A standard approach to implementing this would involve measuring the geometric distance between each of the points within a sculpture and plotting these distances into a histogram with the final measure being set as the Chi-Squared distance between each pair of histograms. The higher the resulting value the more different the sculptures would be.

However, as noted in [5] this process has limitations and the signature generated by the histogram is not unique enough to determine true dissimilarity. In order to combat this several histograms were created for the sculpture similar to the process used in [24]. The full process was not followed as some of the measures were not appropriate for the sculptures, for example using the geodesic distance did not work well as there was no clear surface or path between the points in the sculpture, instead, another approach was formulated which created a histogram for the X, Y and Z distances individually. All histograms were combined, and the Chi-Squared distance was taken on the resulting sculpture signature.

Once each run of the Genetic Algorithm had been completed, the sculpture with the highest dissimilarity was selected and added into an archive. The archive was used in all future runs of the algorithm, to compare all newly generated items with ensuring that all selected items were visually different from each other. Each run of the Genetic Algorithm maximised the dissimilarity between the population and the archive sculptures for 10 generations and a total of 50 runs were completed, providing a selection of 50 sculptures. Similar to the aesthetic tags, 50 sculptures were deemed to be too many to be included in this experiment and so the 18 most visually different sculptures were selected by the authors, the resulting set of 18 sculptures is shown in Figure 1.

3.3 Environment

The final aspect to address was how to present the sculptures to the participants, viewing the 3D artwork on a 2D monitor would be limiting and potentially affect the results considering that the environment is an important aspect of aesthetic appeal [8], [35], [38]. To avoid these limitations VR was chosen, and a custom VR environment was created which allowed the 3D artwork to be viewed within a 3D environment. In addition, VR provides additional benefits which can be utilised to further enhance the user's experience for viewing the artwork.

The concept of using VR environments for viewing artwork is not new and an ever-increasing number of art projects have utilised VR as a medium such as Osmose [12], Chalkroom [3], A Show of Kindness [11] and CAVE [21]. As well as its increasing use within the art world, VR is also becoming a more popular choice for running experiments [15], [17] due, in part, to its ability to explicitly control the environment a participant is placed in and the ability to enhance the art viewing experience.

The VR environment, utilised for this experiment, placed the user within a room resembling a warehouse art gallery with the sculptures placed on a platform along the walls, see Figure 2, the centre of the room formed the main "play area" giving the user a large space where they were able to inspect, move, group, and assign aesthetic tags to each of the sculptures.

Four main interaction techniques were implemented to allow the user to interact with the environment: Translation, Flying, Scaling and Rotation. The interaction techniques used both available controllers to allow the performing of multiple interactions at a single time, for example flying whilst moving a sculpture. The translation interaction technique allowed the user to move a sculpture from one position to another along the floor, flying allowed the user to travel around the environment moving the user forward or backwards at a consistent rate. The movement speed was chosen to avoid potential motion sickness and to allow the user to accurately perform multiple actions whilst flying. Rotation allowed the user to twist the sculpture providing a shortcut to allow a sculpture to be seen from all sides without having to specifically fly around each side. This encouraged the users to view the sculptures from each angle with minimal effort, meaning they could make an informed decision about the descriptive tags. The final interaction technique

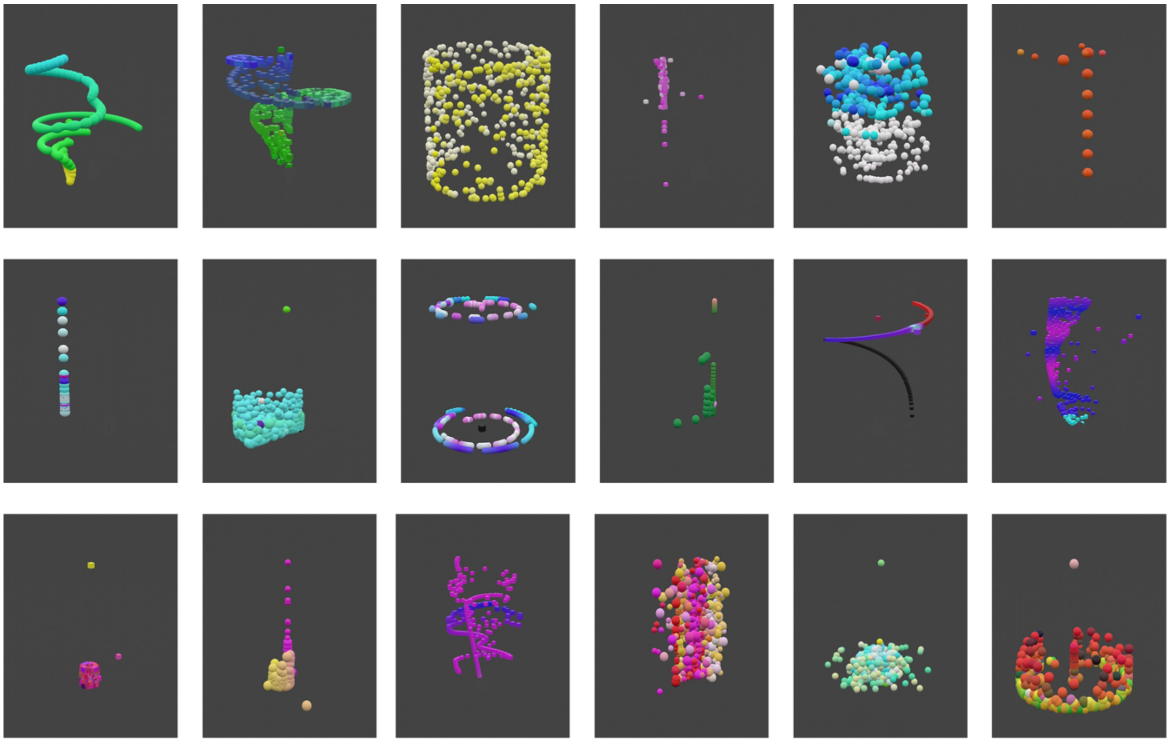

Figure 1: Sculptures available for participants to judge.

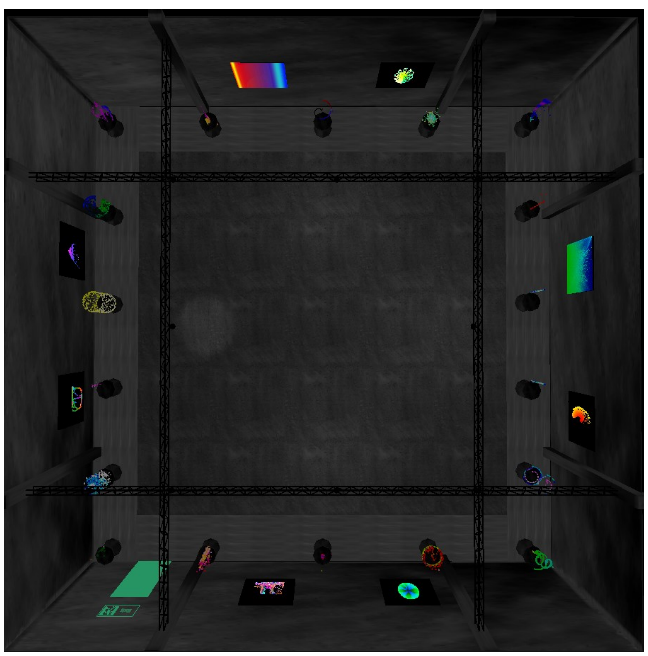

Figure 2: Top-down view showing sculpture starting positions and main play area.

was scaling, this allowed the user to scale themselves down to a tenth of their original size and be placed at the base of a specific sculpture. The users were able to then fly around the sculptures providing a novel experience, uniquely available within a Virtual Reality environment to explore the sculpture from a new perspective, travelling in and out of any part of the sculpture. This afforded the ability to inspect all aspects of the sculpture closely and helped to provide the user with an exciting VR experience.

The main task involved the user choosing their own groups from the provided sculptures based on the minimal instruction of grouping similar sculptures together, with the number and size of the groups being determined by the participant. Detecting the groups was handled by implementing the mean shift clustering algorithm [14], which ran in the background, identifying the clusters the user had placed the sculptures within, allowing the real-time detection of any number of groups formed of any number of sculptures. Each cluster was visually indicated to the user by highlighting the floor around each group (Figure 3), allowing the group to be amended by adding or removing sculptures as necessary. As soon as a group was detected, tags could be chosen from the available list and applied to all sculptures within that group.

3.4 Process

The experiment was designed to be run entirely remotely, using the Prolific platform as the main source of the participants. Participation involved three distinct stages, initially, the users accessed a registration website where they were asked to enter their Prolific Id, read through and agree to the experiment information and consent details. Once this process was complete the user was provided

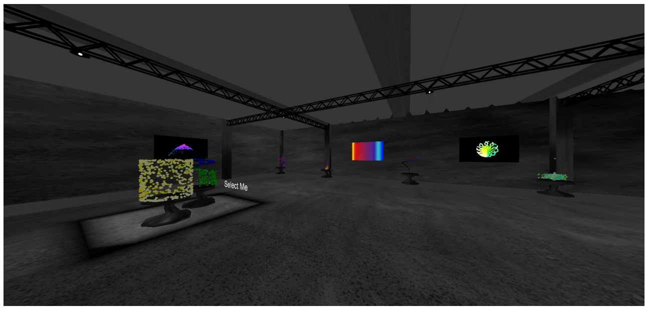

Figure 3: Environment with a detected region.

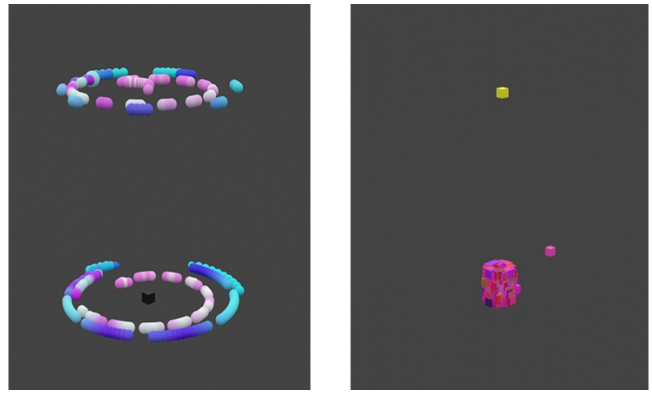

Figure 4: Selection of sculptures with average rating. Highest rated sculpture shown (a). Lowest rated sculpture shown to participants to participants (b).

with a unique code and instructions on how to download the VR environment from the Oculus Quest store[5].

Once the VR app had been downloaded and started, the user was asked to enter their unique code which, after being validated, allowed the anonymous logging of the data collected within the environment. The next stage, comprised of five training tasks, enabled the users to independently learn the four control mechanisms implemented within the environment and learn how to group sculptures and assign tags to the group. This was achieved by providing a task for each technique, such as collecting pellets, moving blocks into a goal, and scaling to collect a key to open a door. The training section of the environment was critical due to the remote nature of running the experiment and so training videos and descriptions were available to the user to help them learn the techniques without anyone being present.

Once the training had been completed, the user completed the main experiment task and was shown the 18 sculptures placed around the edge of the room. The participants were given as long as they wanted to explore and interact with the sculptures whilst creating groups and assigning tags. This approach was chosen as it allowed the participant to change their mind about earlier groupings based on their changing opinions on the sculptures and the tags being assigned.

Once the user had confirmed they had finished creating the groups and assigning the tags, they were prompted to return to the registration website where they were requested to complete a final questionnaire asking questions relating to their experience. Upon completion of the questionnaire, the participants were given the uninstall instructions and a completion code which allowed them to mark their submission on the Prolific platform as complete and collect the payment for participation, which was set as £5 per participant based on the estimated 30 minutes taken for the entire process.

4 RESULTS

To determine whether the chosen aesthetic terms formed any part of the aesthetic judgement of the sculptures and if they did, the extent the terms contributed and whether these terms had a positive or a negative impact on how the sculptures were judged, a large amount of data was collected. This included the participant's position, which was updated every 300ms, the sculptures within their gaze at the same interval, the groups and tags the participant assigned as well as the time the participant spent in each stage. The exit questionnaire collected details about the art expertise of the individual, a rating for each sculpture, from zero to ten, and other general pieces of feedback on the process such as whether the participant experienced any motion sickness and whether any other terms may be appropriate to describe the presented sculptures.

A total of 37 participants completed the experiment (10 female) with a mean age was 30, who spent an average of around 26 minutes within the VR environment. Across all participants, there was a low amount of art expertise as shown by the mode responses provided across all three collected metrics: how often the participant visited an art museum (Rarely), how often an art book was read (Never) and how often the participant practised any form of art (Never). This indicates that the data collected would not be influenced by a high level of expertise allowing the data to be considered more generalised, in terms of art expertise.

The presented sculptures represented a wide range of being aesthetically pleasing to the participants with average ratings of the sculptures ranging from 1.8 (Figure 4b) to 7.3 (Figure 4a).

Table 3 shows the summarised data about each aesthetic tag, intriguingly, it can be seen that Dynamic, Curved, Interesting and Connected terms had the most assignments on average, an early indication of the application difficulty for the tags. One important result is that all available tags were applied to the sculptures suggesting that the chosen tags were suitable for describing the presented sculptures. Dynamic was also the most consistently applied tag, indicating that the sculptures presented exhibited a high level of dynamism (Table 4). Dynamic along with curved, interesting, and connected form some of the main aspects which contribute to the aesthetic judgement of 3D sculptures.

Table 4 also shows an indication of which tags had a suitable level of discrimination and application difficulty to be considered for further analysis, by having a high level of discrimination and a medium level of application difficulty. The tags which fell within the acceptable criteria are highlighted in the table. Dynamic was the most popular tag, due to this, the application difficulty was too low to produce reliable information. Instead, the tags which

[5]https://www.oculus.com/experiences/quest/4632259860201837

Table 3: Total applications (T), How many participants used the tags (P) and the average number of times the tag was applied to a sculpture (A).

Tag	T	P	A
Dynamic	182	30	4.92
Curved	172	31	4.65
Interesting	170	28	4.59
Connected	132	23	3.57
Simple	124	28	3.35
Bright	124	25	3.35
Calm	111	29	3.00
Busy	103	24	2.78
Cold	103	22	2.78
Unnatural	102	23	2.76
Ordered	97	20	2.62
Loud	97	21	2.62
Disordered	97	25	2.62
Complex	95	20	2.57
Boring	95	19	2.57
Warm	86	24	2.32
Natural	77	22	2.08
Static	75	18	2.03
Friendly	69	18	1.86
Separate	64	13	1.73
Original	62	14	1.68
Angular	60	18	1.62
Unfriendly	52	12	1.41
Quiet	51	17	1.38
Sophisticated	51	16	1.38
Unrefined	42	12	1.14
Unoriginal	35	11	0.95
Drab	27	10	0.73

fell within the criteria specified in Section 2 were connected, busy, interesting, ordered, complex, angular, friendly, and calm. The final column shows the positivity rating for each of the tags and all of the terms which met the outlined criteria had a positive influence on the judgement of the sculptures. None of these were the most positively associated terms, however, with this accolade being applied to curved, with simple being the most negatively associated term.

5 DISCUSSION

The collected data raised some interesting patterns, all the available tags were used, with the lowest number of participants being drab which was applied by 10 participants, this is a good indication that the selected terms were fit for purpose and successfully described the generated sculptures, the tags were also mainly rated as positive.

One important aspect of this data is that the details were collected from participants who had a limited amount of art expertise, this potentially helps to explain some of the results that were obtained. As having a high level of art expertise can influence aesthetic judgement the data collected in this experiment can be considered less biased in this respect. This potentially explains the difference in which terms were considered the most important to achieving a positive aesthetic rating over commonly used aspects such as ordered or complex.

Table 4: Consistency (C), Endorsement (E), Difficult of application (Diff), Discrimination (Disc) and positivity rating for each tag, highlighted items represent tags which warrant further investigation.

	C	E	Diff	Disc	Positivity
Dynamic	0.32	5.78	0.68	0.48	0.22
Curved	0.26	4.67	0.74	0.77	0.31
Simple	0.23	4.17	0.77	-0.47	-0.27
Connected	0.23	4.06	0.77	0.81	0.18
Busy	0.18	3.22	0.82	0.42	0.07
Interesting	0.18	3.17	0.82	0.73	0.11
Ordered	0.17	3.00	0.83	0.74	0.11
Complex	0.16	2.94	0.84	0.67	0.13
Bright	0.16	2.83	0.84	0.71	0.08
Disordered	0.15	2.67	0.85	-0.36	-0.08
Unnatural	0.15	2.61	0.86	-0.07	0.00
Static	0.14	2.50	0.86	-0.31	-0.08
Angular	0.12	2.11	0.88	0.36	0.07
Friendly	0.12	2.17	0.88	0.59	0.10
Calm	0.12	2.22	0.88	0.61	0.04
Warm	0.12	2.17	0.88	0.00	0.08
Unfriendly	0.12	2.11	0.88	-0.45	-0.04
Cold	0.11	2.06	0.89	0.10	-0.03
Boring	0.11	2.00	0.89	-0.73	-0.17
Loud	0.11	2.00	0.89	-0.27	-0.01
Separate	0.10	1.89	0.90	-0.48	-0.05
Quiet	0.08	1.39	0.92	-0.07	-0.03
Original	0.08	1.39	0.92	0.43	0.05
Natural	0.08	1.50	0.92	0.49	0.09
Sophisticated	0.07	1.28	0.93	0.27	0.04
Drab	0.06	1.00	0.94	-0.81	-0.11
Unrefined	0.04	0.78	0.96	-0.52	-0.08
Unoriginal	0.03	0.61	0.97	-0.61	-0.08

The positivity or negativity ratings for the terms are mainly intuitive, for example, boring sculptures were rated less highly than interesting sculptures and generally, they corresponded to the antonym relationship between terms e.g. Bright and Drab. However, a few anomalous aspects were identified, for example, whilst natural related to positive ratings, unnatural was neutral, which potentially questions the use of some measures which look at the naturalness of artwork such as the fractal dimension or Benford's law, where having lower ratings on these scales may not be too detrimental. Similarly, the contribution of complexity is positive, however, the positive impact of items being complex is not as significant as the negative contribution if an item is considered simple. This indicates that while complexity is an important aspect of aesthetic judgement, it is more important, when trying to create aesthetically pleasing items, to ensure that items are not simple rather than specifically creating complex items. This potentially affects how items should be generated especially within a Computational Creativity context, where instead of trying to maximise the positive influence of complexity, it should be more important for systems to minimise the negative aspect of simplicity.

The subjectivity of the process is exemplified by the contribution of the curved/angular antonym pair of terms, as both are shown to contribute positively to the judgement, albeit with angular being less positive than curved. Inspecting the sculptures these terms were applied to shows that both terms were applied to the majority of the same sculptures indicating that even though these terms are semantically opposite, both can be displayed within the same sculpture and have a positive effect on the overall rating.

The stalwart formal measures of complexity and order also feature as important aspects, backing up existing research by showcasing that they contribute positively to aesthetic judgement, however, despite their ubiquity, they are not the most positive items which contribute to aesthetic judgement. Terms like curved, dynamic, and connected had a higher positive impact on aesthetic judgement, indicating that in the process of auto-generating aesthetically pleasing items, these terms should be considered more often than complex and ordered.

One final aspect shown from the results is that negative tags seem to be more difficult to describe than positive items. The average difficulty for positive tags was 0.84 whereas for negative tags it was 0.89, which may indicate that negative aspects are less easy to visually distinguish, at least within the sculptures used within this project.

6 CONCLUSION

In this paper, a novel VR environment was presented which allowed the extraction of individual terms that contributed to the aesthetic judgement of 3D sculptures providing a wealth of data about how the judgement is made, along with terms which contribute both positively and negatively to the overall aesthetic rating of each sculpture.

The overall level of expertise of the participants helps the collected data be less biased, due to the influence expertise can have on aesthetic judgement. Successfully collecting data which has been provided by participants with limited expertise in artwork indicates the success the presented approach has in determining which aspects contribute to aesthetic judgement. This approach can be utilised to help overcome, at least partially, the difficulty of describing which aspects contribute to aesthetic judgement, helping to remove this barrier with modelling aesthetic judgement.

The approach also allowed the collection of data which can be used to help guide the auto-generation of artwork, where common aspects e.g. order and complexity, can be considered less important and therefore be less important when attempting to create aesthetically pleasing artwork.

By following set criteria, based on the results obtained through item analysis, potential tags were established which would make good candidates for further investigation. Other items would potentially be good criteria for investigation however they would need a wider set of sculptures to be presented before enough data could be collected to reliably determine the attributes causing their application. The established criteria considered the application difficulty and discrimination of each tag, where a medium level of difficulty and a high level of discrimination was required to select a tag for further consideration. The collected data led to nine tags being identified: connected, busy, interesting, ordered, complex, bright, angular, friendly, and calm, which can be used as a basis for formalisation.

However, there were limitations with this process, using a small set of geometric shapes to generate the sculptures, only spheres and cubes, may have influenced how easily terms such as curved or angular were applied. Whilst care was taken to try and present a wide range of terms for use, by only presenting a set of 18 terms, the process placed restrictions on how the presented items could be judged. The result suggests that the terms were suitable however it is possible that other terms could have a greater positive or negative impact than the ones identified. In addition, the restriction also means that the process cannot represent a full aesthetic judgement, only a small portion of it. The participants were given the opportunity to suggest other terms, which could help to rectify these limitations, however, there was no agreement, with each participant suggesting different terms, a further indication of how subjective the process of judging artwork is and how successfully this approach of extracting aspects was. The small amount of sculptures also represents a limitation with this system, again even though effort was taken to make the sculptures as different from each other as possible, only a small selection was presented, which may not have exhibited all of the provided terms in equal amounts.

Overall, several important and relatively unconsidered aesthetic aspects were identified through utilising item analysis to determine how well the aesthetic tags described the presented sculptures. This provided a wealth of information on how the tags were applied and the positive and negative impact of the terms but also provided 3D models which exhibit the properties of each term, helping to identify some of the many aspects which contribute to the aesthetic judgement of 3D sculptures.

6.1 Future Work

The data collected in this experiment opens up multiple avenues for further research, such as looking at the difference between the positive or negative interpretation of the aesthetic terms. However, the main approach to be investigated by the authors is attempting to create a formalisation of the identified suitable tags to allow their introduction to auto-generating 3D sculptures. The success of these formal measures will be determined before combining all this information to generate sculptures based on individual preferences.

The approach presented in this experiment can also be amended to extend the amount of information that is collected, for example, it could be applied to 2D images and then compared to the 3D items to investigate what differences exist between the judgement of the two types of items. A potentially wider array of terms and sculptures could be presented which would further extend the aspects which can be identified along with further understanding of the interaction between the terms and whether this affects the positive or negative connotations.

REFERENCES

[1] E. Acebo and M. Sbert. 2005. Benford's law for natural and synthetic images. In First Eurographics conference on Computational Aesthetics in Graphics, Visualization and Imaging.

[2] Mohammad Majid al Rifaie, Anna Ursyn, Robert Zimmer, and Mohammad Ali Javaheri Javid. 2017. On symmetry, aesthetics and quantifying symmetrical complexity. In *International Conference on Evolutionary and Biologically Inspired Music and Art*. Springer, 17–32.

[3] Laurie Anderson and Hsin-Chien Huang. 2017. Chalkroom. [VR Animation].

[4] Steve Bergen and Brian J Ross. 2013. Aesthetic 3D model evolution. *Genetic Programming and Evolvable Machines* 14, 3 (2013), 339–367.

[5] Silvia Biasotti, Andrea Cerri, Alex Bronstein, and Michael Bronstein. 2016. Recent trends, applications, and perspectives in 3d shape similarity assessment. In *Computer graphics forum*, Vol. 35. Wiley Online Library, 87–119.

[6] Ado Abdu Bichi. 2016. Classical Test Theory: an introduction to linear modeling approach to test and item analysis. *International Journal for Social Studies* 2, 9 (2016), 27–33.

[7] George David Birkhoff. 1933. *Aesthetic measure*. Cambridge, Mass.

[8] Thenille Braun Janzen, Bruna de Oliveira, Giulia Ventorim Ferreira, João R Sato, Claudia Feitosa-Santana, and Patrícia Vanzella. 2023. The effect of background music on the aesthetic experience of a visual artwork in a naturalistic environment. *Psychology of Music* 51, 1 (2023), 16–32.

[9] Daniel G Brown and Dan Ventura. 2022. Ethics, Aesthetics and Computational Creativity. In *ICCC*.

[10] Jonathan Byrne, Erik Hemberg, Michael O'Neill, and Anthony Brabazon. 2013. A methodology for user directed search in evolutionary design. *Genetic Programming and Evolvable Machines* 14, 3 (2013), 287–314.

[11] Jeremy Cowles, Peter Chan, Tory Voight, and Isabel Parkinson. 2018. A Show of Kindness. In *ACM SIGGRAPH 2018 Virtual, Augmented, and Mixed Reality* (Vancouver, British Columbia, Canada) *(SIGGRAPH '18)*. Association for Computing Machinery, New York, NY, USA, Article 2, 1 page. https://doi.org/10.1145/3226552.3226572

[12] Char Davies. 1995. Osmose: Notes on Being in Immersive Virtual Space. In *ISEA*'95.

[13] Eelco den Heijer and Agoston E Eiben. 2010. Comparing aesthetic measures for evolutionary art. In *European Conference on the Applications of Evolutionary Computation*. Springer, 311–320.

[14] Konstantinos G Derpanis. 2005. Mean shift clustering. *Lecture Notes 32* (2005), 1–4.

[15] Edward Easton, Ulysses Bernardet, and Aniko Ekart. 2019. Tired of choosing? Just add structure and virtual reality. In *Computational Intelligence in Music, Sound, Art and Design*: 8th International Conference, EvoMUSART 2019, Held as Part of EvoStar 2019, Leipzig, Germany, April 24–26, 2019, Proceedings 8. Springer, 142–155.

[16] Edward Easton, Anikó Ekárt, and Ulysses Bernardet. 2022. Axial Generation: Mixing Colour and Shapes to Automatically Form Diverse Digital Sculptures. *SN Computer Science* 3, 6 (2022), 505.

[17] Doga Gulhan, Szonya Durant, and Johannes M Zanker. 2022. Aesthetic judgments of 3D arts in virtual reality and online settings. *Virtual Reality* (2022), 1–17.

[18] Gregor U Hayn-Leichsenring and Anjan Chatterjee. 2019. Colliding terminological systems—Immanuel Kant and contemporary empirical aesthetics. *Empirical Studies of the Arts* 37, 2 (2019), 197–219.

[19] Bryan Hollingsworth and Jacob Schrum. 2019. Infinite art gallery: a game world of interactively evolved artwork. In *2019 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 474–481.

[20] Colin G Johnson, Jon McCormack, Iria Santos, and Juan Romero. 2019. Understanding aesthetics and fitness measures in evolutionary art systems. *Complexity* 2019 (2019).

[21] Kris Layng, Ken Perlin, Sebastian Herscher, Corinne Brenner, and Thomas Meduri. 2019. CAVE: Making Collective Virtual Narrative: Best Paper Award. *Leonardo* 52, 4 (2019), 349–356.

[22] Helmut Leder and Marcos Nadal. 2014. Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode–Developments and challenges in empirical aesthetics. *British journal of psychology* 105, 4 (2014), 443–464.

[23] Helmut Leder, Pablo PL Tinio, David Brieber, Tonio Kröner, Thomas Jacobsen, and Raphael Rosenberg. 2019. Symmetry is not a universal law of beauty. *Empirical Studies of the Arts* 37, 1 (2019), 104–114.

[24] Mona Mahmoudi and Guillermo Sapiro. 2009. Three-dimensional point cloud recognition via distributions of geometric distances. *Graphical Models* 71, 1 (2009), 22–31.

[25] Kresimir Matkovic, László Neumann, Attila Neumann, Thomas Psik, and Werner Purgathofer. 2005. Global contrast factor-a new approach to image contrast. *Computational Aesthetics* 2005 (2005), 159–168.

[26] Jon McCormack, Camilo Cruz Gambardella, and Andy Lomas. 2021. The enigma of complexity. In Artificial Intelligence in Music, Sound, Art and Design: 10th International Conference, EvoMUSART 2021, Held as Part of EvoStar 2021, Virtual Event, April 7–9, 2021, Proceedings 10. Springer, 203–217.

[27] Jon McCormack and Andy Lomas. 2020. Understanding Aesthetic Evaluation using Deep Learning. In *International Conference on Computational Intelligence in Music, Sound, Art and Design (Part of EvoStar)*. Springer, 118–133.

[28] James McDermott. 2012. Graph grammars as a representation for interactive evolutionary 3d design. In *International Conference on Evolutionary and Biologically Inspired Music and Art*. Springer, 199–210.

[29] James McDermott, John Mark Swafford, Martin Hemberg, Jonathan Byrne, Erik Hemberg, Michael Fenton, Ciaran McNally, Elizabeth Shotton, and Michael O'Neill. 2012. String-rewriting grammars for evolutionary architectural design. *Environment and Planning B: Planning and Design* 39, 4 (2012), 713–731.

[30] Luis Carlos Pereira Monteiro, Victória Elmira Ferreira do Nascimento, Amanda Carvalho da Silva, Ana Catarina Miranda, Givago Silva Souza, and Rachel Coelho Ripardo. 2022. The role of art expertise and symmetry on facial aesthetic preferences. *Symmetry* 14, 2 (2022), 423.

[31] Manuel Muehlbauer, Jane Burry, and Andy Song. 2017. Automated shape design by grammatical evolution. In *International Conference on Evolutionary and Biologically Inspired Music and Art*. Springer, 217–229.

[32] Miguel Nicolau and Dan Costelloe. 2011. Using grammatical evolution to parameterise interactive 3D image generation. In *European Conference on the Applications of Evolutionary Computation*. Springer, 374–383.

[33] Michael O'Neill, James McDermott, John Mark Swafford, Jonathan Byrne, Erik Hemberg, Anthony Brabazon, Elizabeth Shotton, Ciaran McNally, and Martin Hemberg. 2010. Evolutionary design using grammatical evolution and shape grammars: Designing a shelter. *International Journal of Design Engineering* 3, 1 (2010), 4–24.

[34] Una-May O'Reilly and Martin Hemberg. 2007. Integrating generative growth and evolutionary computation for form exploration. *Genetic Programming and Evolvable Machines* 8, 2 (2007), 163–186.

[35] Matthew Pelowski, Michael Forster, Pablo PL Tinio, Maria Scholl, and Helmut Leder. 2017. Beyond the lab: an examination of key factors influencing interaction with 'real'and museum-based art. *Psychology of Aesthetics, Creativity, and the Arts* 11, 3 (2017), 245.

[36] Kennedy Quaigrain and Ato Kwamina Arhin. 2017. Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education* 4, 1 (2017), 1301013.

[37] Frank Sibley. 1959. Aesthetic concepts. *The philosophical review* 68, 4 (1959), 421–450.

[38] Eva Specker, Pablo PL Tinio, and Michiel van Elk. 2017. Do you see what I see? An investigation of the aesthetic experience in the laboratory and museum. *Psychology of Aesthetics, Creativity, and the Arts* 11, 3 (2017), 265.

[39] Pablo PL Tinio and Helmut Leder. 2009. Just how stable are stable aesthetic features? Symmetry, complexity, and the jaws of massive familiarization. *Acta Psychologica* 130, 3 (2009), 241–250.

[40] Sibel Toksöz and Ayşe Ertunç. 2017. Item analysis of a multiple-choice exam. *Advances in Language and Literary Studies* 8, 6 (2017), 141–146.

[41] Dan Ventura and Darin Gates. 2018. Ethics as Aesthetic: A Computational Creativity Approach to Ethical Behavior. In *ICCC*. 185–191.