Architecture and Planning Journal (APJ)

Volume 28 Issue 3 *ASCAAD 2022 - Architecture in the Age of the Metaverse – Opportunities and Potentials*
ISSN: 2789-8547

Article 43

March 2023

MINDFUL SPACE IN SENTENCES - A DATASET OF VIRTUAL EMOTIONS FOR NATURAL LANGUAGE CLASSIFICATION

HAN TU
Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, USA, hantu@mit.edu

CHUNFENG YANG
Harvard Graduate School of Design, 48 Quincy St, Cambridge, Massachusetts, USA, chunfengyang@gsd.harvard.edu

MINDFUL SPACE IN SENTENCES - A DATASET OF VIRTUAL EMOTIONS FOR NATURAL LANGUAGE CLASSIFICATION

Abstract

Spatial emotions have played a critical role in visual-spatial environmental assessment, which can be assessed using bio-sensors and language description. However, information on virtual spatial emotion assessment with objective emotion labels and natural language processing (NLP) is insufficient in literature. Thus, designers' ability to assess spatial design quantitatively and cost effectively is limited before the design is finalized. This research measures the emotions expressed using electroencephalograms (EEGs) and descriptions in virtual reality (VR) spaces with different parameters. First, 26 subjects experienced 10 designed virtual spaces with a VR headset (Quest 2 device) corresponding to the different space parameters of shape, height, width, and length. Simultaneously, the EEG measured the emotions of the subjects using four electrodes and the five brain waves. Second, two labels – calm and active – were produced using EEGs to describe these virtual reality spaces. Last, this labeled emotion dataset compared the differences among the virtual spaces, human feelings, and the language description of the participants in the VR spatial experience. Experimental results show that the parameter changes of VR spaces can arouse significant fluctuations in the five brain waves. The EEG brain wave signals, in turn, can label the virtual rooms with calm and active emotions. Specifically, in terms of VR spaces and emotions, the experiments find that more relative spatial height results in less active emotions, while round spaces arouse calmness in the human brain waves. Moreover, the precise connection among VR spaces, brain waves in emotion, and languages still needs further research. This research attempts to offer a useful emotion measurement tool in virtual architectural design and description using EEGs. This research identifies potentials for future applications combining physiological metrics and AI methods, i.e., machine learning for synthetic design generation and evaluation.

Keywords

Virtual reality (VR), Spatial Perception, Spatial assessment, Electroencephalogram (EEG), Natural Language Processing (NLP)

MINDFUL SPACE IN SENTENCES
A DATASET OF VIRTUAL EMOTIONS FOR NATURAL LANGUAGE CLASSIFICATION

HAN TU[1] AND CHUNFENG YANG[2]

[1] Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts, USA
hantu@mit.edu

[2] Harvard Graduate School of Design, 48 Quincy St, Cambridge, Massachusetts, USA
chunfengyang@gsd.harvard.edu

ABSTRACT

Spatial emotions have played a critical role in visual-spatial environmental assessment, which can be assessed using bio-sensors and language description. However, information on virtual spatial emotion assessment with objective emotion labels and natural language processing (NLP) is insufficient in literature. Thus, designers' ability to assess spatial design quantitatively and cost effectively is limited before the design is finalized. This research measures the emotions expressed using electroencephalograms (EEGs) and descriptions in virtual reality (VR) spaces with different parameters. First, 26 subjects experienced 10 designed virtual spaces with a VR headset (Quest 2 device) corresponding to the different space parameters of shape, height, width, and length. Simultaneously, the EEG measured the emotions of the subjects using four electrodes and the five brain waves. Second, two labels – calm and active – were produced using EEGs to describe these virtual reality spaces. Last, this labeled emotion dataset compared the differences among the virtual spaces, human feelings, and the language description of the participants in the VR spatial experience. Experimental results show that the parameter changes of VR spaces can arouse significant fluctuations in the five brain waves. The EEG brain wave signals, in turn, can label the virtual rooms with calm and active emotions. Specifically, in terms of VR spaces and emotions, the experiments find that more relative spatial height results in less active emotions, while round spaces arouse calmness in the human brain waves. Moreover, the precise connection among VR spaces, brain waves in emotion, and languages still needs further research. This research attempts to offer a useful emotion measurement tool in virtual architectural design and description using EEGs. This research identifies potentials for future applications combining physiological metrics and AI methods, i.e., machine learning for synthetic design generation and evaluation.

Keywords: Virtual reality (VR), Spatial Perception, Spatial assessment, Electroencephalogram (EEG), Natural Language Processing (NLP)

ملخص

لقد طالما لعبت العواطف المكانية دورًا مهمًا في التقييم البيئي البصري المكاني، والذي يمكن تقييمه باستخدام أجهزة الاستشعار الحيوية ووصف اللغة. ومع ذلك، فإن المعلومات المتعلقة بتقييم المشاعر المكانية الافتراضية مع تسميات العاطفة الموضوعية وتقنيات معالجة اللغة الطبيعية غير كافية في الأدبيات ذات الصلة. وبالتالي ، فإن قدرة المصممين على تقييم التصميم المكاني من الناحية الكمية وفعالية التكلفة محدودة قبل الانتهاء من التصميم. يقيس هذا البحث المشاعر التي يتم التعبير عنها باستخدام مخطط كهربية الدماغ (EEGs) وتوصيفات في مساحات الواقع الافتراضي طبقا لمجموعة مختلفة من المعايير. أولاً اختبر ٢٦ شخصًا ١٠ مساحات افتراضية مصممة باستخدام سماعة الرأس الافتراضية (جهاز Quest 2) تتوافق مع مجموعة من متغيرات الفراغ مثل الشكل والارتفاع والعرض والطول. في الوقت ذاته، تم قياس عواطف المستخدمين باستخدام أربعة أقطاب كهربائية وموجات الدماغ الخمسة عن طريق مخطط كهربية الدماغ (EEG). ثانيًا، تم إنتاج ملصقين - هادئ ونشط - باستخدام تخطيط كهربية الدماغ لوصف فراغات الواقع الافتراضي تلك. أخيرًا، تمت مقارنة الاختلافات بين الفراغات الافتراضية والمشاعر الإنسانية والوصف اللغوي للمشاركين في التجربة المكانية للواقع الافتراضي عن طريق مجموعات بيانات المشاعر. تظهر النتائج التجريبية أن التغير فى المتغيرات لفراغات الواقع الافتراضي يمكن أن تثير تقلبات كبيرة في موجات الدماغ الخمس. يمكن لإشارات موجة الدماغ EEG ، بدورها تسمية الغرف الافتراضية بمشاعر هادئة ونشطة. على وجه التحديد، فيما يتعلق بفراغات الواقع الافتراضي والعواطف، وجدت التجارب أن الارتفاع المكاني النسبي يؤدي إلى عواطف أقل نشاطًا، بينما تثير الفراغات المستديرة الهدوء في موجات الدماغ البشري. علاوة على ذلك، لا يزال الاتصال الدقيق بين فراغات الواقع الافتراضي وموجات الدماغ في العاطفة واللغات بحاجة إلى مزيد من البحث. يحاول هذا البحث تقديم أداة مفيدة لقياس المشاعر في التصميم والوصف المعماري الافتراضي باستخدام مخطط كهربية الدماغ. يحدد هذا البحث إمكانات التطبيقات المستقبلية التي تجمع بين المقاييس الفسيولوجية وطرق الذكاء الاصطناعي، أي التعلم الآلي لتوليد التصميم التركيبي وتقييمه.

الكلمات المفتاحية: الواقع الافتراضى، الإدراك الفراغى، التقييم الفراغى، مخطط كهربية الدماغ، تقنيات معالجة اللغة الطبيعية.

1. INTRODUCTION

Design evaluation in the real-world costs much before the design is finalized. Moreover, it's expensive to incorporate user feedback with sentiments and description post-design due to the individual differences. Quantifying and explaining user experience are important to detect patterns to make generalizations in design assessment.

The immersive experience of virtual reality (VR) offers a new way to measure the impact of pure visual experiences (El Beheiry et al., 2019) with lower cost before the design is finalized. Researchers can simulate intricate real-life spaces in VR to investigate the emotion feedback that subjects experience inside the spatial settings. Although there are many VR and electroencephalograms (EEGs) studies (Shemesh et al., 2021; Suhaimi et al., 2020), their spaces lack the elements of everyday architectural spaces, such as windows and doors (Shemesh et al., 2017), and the language interpretation of the spaces by the participating experimenters (Hu et al., 2019). Our study hypothesizes that we can quantify people's descriptions and explanations of spaces by analyzing their sentences with emotional labels.

The immersive spaces in this research consist of 10 rooms with variable design parameters of shape, height, width, and length to stimulate emotional experiences of spaces that the users experience during everyday life. In the experimental study, 26 participants describe the virtual spaces of different parameters and explain their reactions in sentences to create a spatial dataset. The dataset is built in 3 dimensions of spatial parameters, sentences, and emotional labels to assess the spatial emotions in everyday language. Our study aims to assess how design parameters influence linguistic and emotional responses to virtual spaces. Our dataset assists architects' decision-making processes in facilitating more accessible and accurate labeling of description and emotion reactions to spaces.

2. RELATED WORK

Virtual reality (VR) technology has been used as an architectural assessment tool to assist in decision-making for unbuilt spaces, as well as psychological, educational, and recreational design tools. Measurements of architectural spaces in VR involve measuring quantitative spatial parameters and qualitative spatial features. Quantitative spatial measures are mostly related to size, proportion, scale, and distance perceptions (Shemesh et al., 2021), while qualitative features are the environmental aspects of spaces, such as openness and closeness (Ergan et al., 2018b, 2018a). Assessments of architectural spaces in VR involve the feedback from user physiological data and language explanation, such as descriptions and questionnaires.

In the assessment of VR spaces, non-architects find it difficult to explain their exact feelings about the visual space. For example, non-architects will describe a space with some sentences, like "it is a space for me to study calmly in," rather than a more straightforward word, like "tranquil" or "unaroused." To quantify emotional reactions from participants in VR spaces, researchers build body biosensor systems, including electroencephalogram (EEG), galvanic skin response (GSR), heart rate in photoplethysmogram (PPG) and eye tracking, to measure and assess basic emotions in VR spatial experiences (Ergan et al., 2018a). Among all the biosensors, EEGs play a role in measuring the emotion of brain waves to measure cognitive emotion and assess architectural design with descriptions (Diemer et al., 2015). In this research, the emotional states of calmness and activeness are measured with EEGs corresponding to the descriptions during the spatial experiences.

Natural language processing (NLP) has been recognized as a critical language tool for statistical analysis and assessment of designed spaces. However, everyday descriptions of the experience with natural language, such as the text sentences from Twitter, seldom result from only visual-spatial experiences. These sentences describe the spaces merged with other elements, such as the temperature, social interactions, and senses. Architecture researchers have difficulty distinguishing whether the experiences in a sentence come from space or from other elements of everyday life, such as food and weather. Existing emotion classifiers do not provide effective guidance to architects in practical applications (Mohammad et al., 2018). Thus, the architecture researchers' ability to evaluate spaces using language is limited. This research aims to build a dataset of natural language mapping from visual-spatial sentence descriptions to visual-spatial emotion labels.

3. METHODOLOGY

This section introduces the parameters of VR models, the apparatus, and the psychophysiological Indices of biosensors (EEG devices), overview of participants, experiment procedures and data analyzing methodology (Figure 1). The methods aim to match the sentence descriptions with the corresponding emotion and spatial parameters using EEG and VR models.

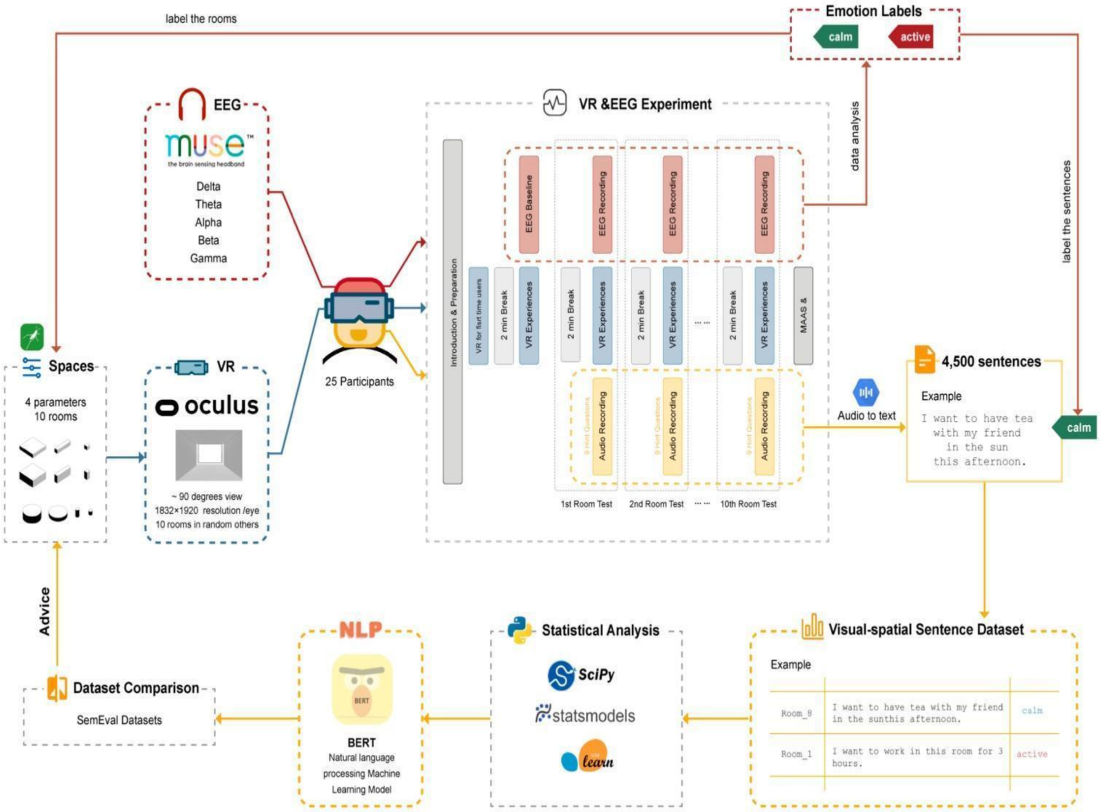


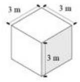
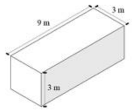
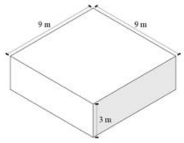
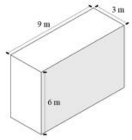
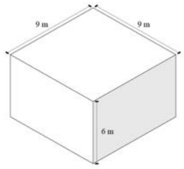
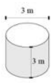
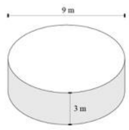
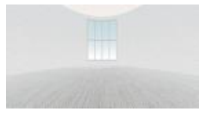
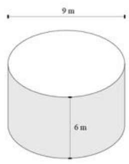
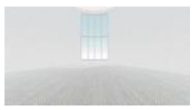
Fig.1: A diagram showing the pipeline of the methodology using EEGs, VR and NLP

3.1. Architectural Spaces and VR Devices

Researchers rendered 10 3D room models in Unity for the VR headset experiences. The parameters of the virtual rooms (Table 1) are the shapes of the rooms (rectangle, and round in diameter of 3 m or 9 m), and the sizes (height in 3 m or 6 m, width in 3 m or 9 m, and length in 3 m or 9 m) of the rooms. Other virtual spatial parameters, such as color, light and the size proportion of the window (0.5 m, 0.5 m, 0.5 m, 0), remain the same. Many of these prototyped rooms are shaped in design strategies as potential spaces for new study, work, and recreation. All of them can play an important role providing calming space for mindfulness.

The participants were also equipped with a wearable VR device – Oculus Quest 2. The Quest 2 offers a VR experience with the field of view in 90 degrees and per eye resolution of 1832×1920. The researchers set up the 10 room models in Quest 2 for VR testing. All these space models are then recognized as calm (mindful) or active (anxious) spaces using electroencephalogram (EEG) devices with a spontaneous effect on human brain waves in five brainwaves: alpha, beta, gamma, theta, and delta.

Table 1: The parameters of the 10 VR rooms

Room No.	Shape	Height	Length	Width	Room Settings	Immersive View
1	Rectangle	3 m	3 m	3 m		
2	Rectangle	3 m	3 m	9 m		
3	Rectangle	3 m	9 m	9 m		
4	Rectangle	6 m	3 m	3 m		
5	Rectangle	6 m	3 m	9 m		
6	Rectangle	6 m	9 m	9 m		
7	Round	3 m	3 m			
8	Round	3 m	9 m			
9	Round	6 m	3 m			
10	Round	6 m	9 m			

3.2. Biosensors

Viewing the VR rooms, the participants were also equipped with a Muse headset – a wearable brain sensing headband. The device is fully portable and can be paired with any tablet or smartphone and operated with the Muse application (Mind Monitor) to record the EEG data. The sensors are placed on the frontal lobes of subjects to detect the five brain waves (Table 2) – delta (1-4 Hz), theta (4-8 Hz), alpha (7.5-13 Hz), beta (13-30 Hz) and gamma (30-44 Hz). The EEG device tracks and records the five brain wave patterns in 4 electrodes of signal collected from frontal areas (TP9, AF7, AF8 and TP10 electrodes). Brain waves data from Muse 2 are absolute band powers, based on the logarithm of the Power Spectral Density (PSD) of the EEG data for each channel. When analyzing the EEG data, we compared the data of five types of oscillations.

Table 2: The five brainwaves' absolute frequency ranges read from the Muse headset

Name	Frequency Range	Description in emotion
Delta	1–4 Hz	Usually indicating the unconscious mind and occurs in deep sleep
Theta	4–8 Hz	Usually indicating the subconscious mind and occurs in sleeping and dreaming
Alpha	7.5–13 Hz	Usually indicating a relaxed mental state yet aware and are correlated with brain activation
Beta	13–30 Hz	Usually indicating active mind state and occurs during intense focused mental activity
Gamma	30–44 Hz	Usually associated with intense brain activity

3.3. Comments from the Participants

Individuals differ in terms of baseline anxieties and arousal levels in reacting to the VR room experiences. The Mindful Attention Awareness Scale (MAAS) can measure mindfulness as an attribute that varies between people (Brown and Ryan, 2003). Thus, MAAS measures the diversity of attention to and awareness of present events and experiences in the VR spaces. The questionnaire includes questions related to the elements of VR spaces and the experience. The question settings of the questionnaire are 5 gradients - almost always, very frequently, something frequently, something infrequently, very infrequently, and almost never, while people cognize and feel the spaces hardly by these numerical values. Moreover, the questionnaire questions are difficult to cover all the spatial cognition and emotional feedback of the participants. To collect more feedback in the experiments, participants expressed their spontaneous spatial feelings through their language. Researchers used audio to record participants' comments on the spaces and questions (Table 3). After converting the audio to text, researchers conduct a natural language processing analysis of the sentences combined with emotion from EEGs.

Table 3: The questions for each participant during the VR test for each room

	Feeling Sentences
Question 1	How would you describe the space in more than two sentences?
Question 2	How would you describe your feelings of the space in more than two sentences?
	Behavioral Intention Sentences
Question 3	What would you like to do in this room right now?
Question 4	What would you like to do in this room for the following 3 hours?
Question 5	What would you like to do in this room with your best friends?
	Spatial Intention Sentences
Question 6	If this room was part of your house, then, which room of your house do you think this room would be and why?
Question 7	If this room was a public room, what kind of function do you think the room would have?
Question 8	If you could name this room, what name or title would you want to choose and why?
Question 9	The previous subject described the room as "a prison"; what do you think about that?

3.4. Experiment Setup

The participants consisted of n = 26 people (16 women, 10 men; mean age = 27, SE = 7.65). They were architecture or non-architecture students and volunteers from higher education institutes, such as MIT and Harvard. 6 of them do not have VR experience before and 20 of them have VR experiences before our test.

The experiment was designed to be independently executed by each of the participants, and each session was recorded individually (Table 4). They were also provided with a VR headset (Oculus Quest 2), Muse headset to measure brain waves, a chair to sit on, and a quiet area to stay for 30 minutes (Figure 2). All participants underwent a short training and a short sample session with the apparatus. They were instructed by the researchers on how to perform the experiment—describe their views or feelings of the VR rooms in sentences following the questions (Table 3) designed to elicit appropriate responses. After this, participants experienced the rooms in a random order, describing their feelings in audio-recorded sentences.

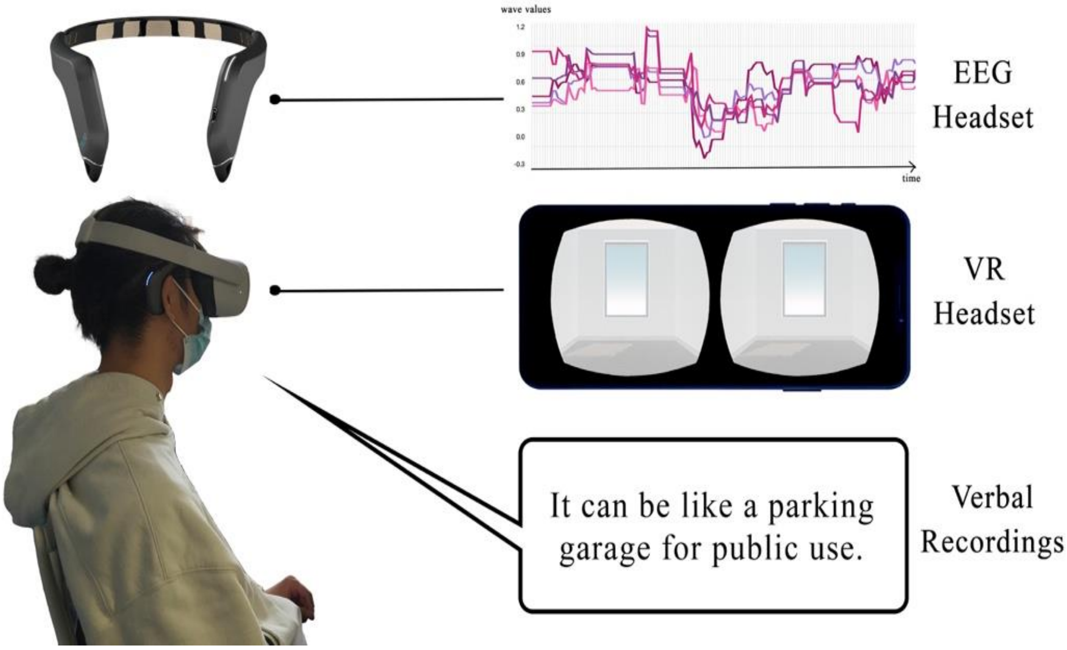


Fig.2: Example of participants' experiments with EEG headset, VR headset, and Verbal recordings

In the VR experiment, connectivity with the Muse headset had to be established with Bluetooth and checked for all four electrodes (TP9, AF7, AF8, TP10) each time. After the proper connection of smartphones with Muse, the baseline phase of the experiment—the relaxation phase—was established. During this time, participants were asked to close their eyes and relax for 30 seconds while their EEG data were recorded by Muse (the baseline phase, was later on subtracted from the signal). Next, during the 60 minutes' sitting, participants looked at the VR rooms without unnecessary movement. Viewing the VR rooms, the participants answered the questions asked by the researchers. Complete recording of sentences at 10 room views (plus the break time of 2 min each for interval) took between 45 and 60 min. The participants filled out the questionnaires after doing the VR experiment.

Table 4: Research design and the phases of the procedures

	Preparation for VR Research
Phase 1	10 VR rooms designed by researchers (Section 3.1); Random choice of rooms' orders and questions for participants to answer (Table 3).
	Introduction for Participants
Phase 2	Introduction to research goals and equipment — instructions for all 26 participants. Presentation of research VR headset including 10 rooms in random order. Test Bluetooth connection—smartphone to Muse EEG device. First-time VR users will have 1 min more to experience the VR headset to overcome the initial excitement of trying VR for the first time. Then, the participants had 30 seconds' rest before the VR Research and Audio Recordings.
	VR Experiments and Audio Recordings
Phase 3	The participants are seated in a room without sound or movement for the VR tests. During the VR spatial test, the participants wear the Muse Headset to record the EEG data simultaneously. The researchers ask the participants the questions (Table 3) on the topics of spatial feeling, behavior intention, and random sentences for each virtual room. The participants answer in sentences to describe each virtual room. The researchers audio-record the sentences to match the sentences and the emotions from EEG to build the dataset. The participants have a 2 min break in the transition of the two VR rooms to avoid emotional influences of the room transition.
	Individual Mindfulness Balanced Questionnaire
Phase 4	The Mindful Attention Awareness Scale questionnaires (Brown and Ryan, 2003) together with other VR experiences related questions measured the individual emotional differences.
	Data Curation
Phase 5	Data preparation as described in the "Statistical Analysis" section of this article.
	Data Analysis
Phase 6	Data analysis as described in the "Results" section of this article.
	Dataset Building and NLP
Phase 7	Dataset preparation and NLP training as described in the "NLP Classification Training" section of this article.

4. RESULTS

Our dataset contains 1,402 sentences from participants' comments with 2 labels and 10 rooms (Table 5). We analyzed our dataset using electroencephalograms (EEGs), statistical tools and natural language processing (NLP) classification models. In this section, we showcase the potential of using our dataset for data-driven analysis of spaces and emotions.

Table 5: Example of our dataset with1,402 sentences, 2 emotion labels and 10 rooms

Room ID	Sentences	Emotion label
6	This is very upright space, and I'm looking at this strike so weird.	1
6	It looks like terrace; I don't know how you attached the texture.	0
3	It's a similar room, I would say, still upright, but seems like the window says is smaller, but it's still like large accordingly relative to the wall that I'm facing. So, I would say it's still like a very, like nice view room.	0
3	I think it's large, I would not say it's huge, but it's a size of like good. I like a small dance hall, but I see is quite low, like a dorm ceiling.	0
8	I feel like I'm in a barrel. So, I've already decided the nickname. Because it's cylinder and very tall. So, make me like a can, and it's weird, because you are there any kind of flaw here, because like this is a curry wall and like the door, as well as the handrail, has some bug.	1
8	It has a curvy like window.	0

4.1. Analysis from the Users' Comments

In terms of the relationship between rooms and emotions, each room can arouse some participants' calm and active emotions. Therefore, it is difficult to label a room with emotion. Moreover, some rooms have distinct emotional features, such as room 1 (shape = rectangle, height = 3m, length = 3m, width = 3m), room 6 (shape = rectangle, height = 6m, length = 9m, width = 9m), and room 5 (shape = rectangle, height = 6m, length = 3m, width = 9m), and some do not have distinct emotional features (Figure 3).

Furthermore, Figure 3 illustrates that rooms have stark individual differences in emotional arousal. Participants may have generally more active emotions (e.g., Participant 9, 10, 15, and 26), or calmer emotions (e.g., Participants 7, 13, and 23) in the virtual reality (VR) environment. Such deviations may come from the deviation of individual spatial cognition and spatial perception, or the individual's different stress responses to VR space during the experiments. For example, Participant 25 said he was excited when speaking and wondering in the VR rooms.

In terms of the relationship between sentences and emotions, the emotional labels of sentences were strongly correlated with participant number and room height parameters. The significant differences in emotions among the participants also reflect the difference in the individual's emotions towards the room. However, for the higher spaces, most participants exhibited calmer emotions (Table 6, -0.20** with significance in the relationship between height and active). Therefore, we expect further analysis between sentences and emotions.

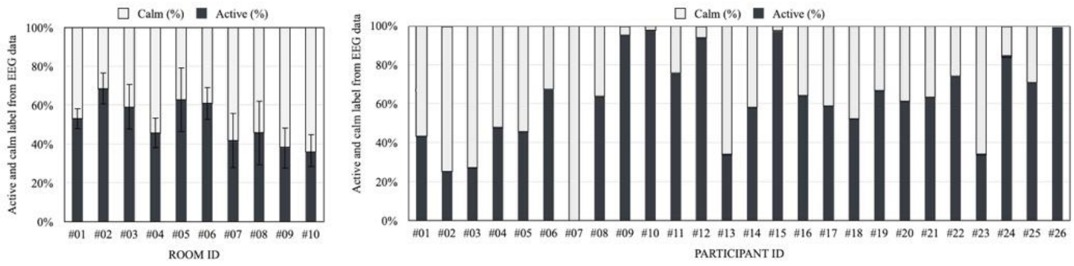


Fig.3: Calm and active distribution by Room ID and Participant ID

Table 6: Correlations between emotion labels of rooms and participant ID

	Participant_ID	Emotion_label
Participant_ID (gender/architects/)	1	**.303****
Room_ID	.051	-.035
Length	**-.101****	-.037
Width	-.009	.076
Height	-.038	**-.020****
Emotion_label	**.303****	1

4.2. EEG Results

To prepare the emotion label data from the EEGs, we carry out additional preprocessing and then statistical analyses using Python programming language (with the libraries SciPy, Statsmodels and Scikit_posthocs).

The EEG signal from the Muse 2 headset is recorded with the use of the Muse Monitor mobile app. The absolute signal undergoes a Fast Fourier Transform (FFT) algorithm in Muse to compute the power spectral density of each frequency range (for instance, alpha, i.e., 9-13 Hz) on each channel. Basically, the signal from each channel shows "how much" of each frequency exists in the interval of one second. The raw data are given on a log scale, in units of Bels, for further analysis. At first, the signal is cleaned in terms of the number of non-informative elements and errors (e.g., missing data, non-numeric values, poor signal quality obtained from a particular electrode, eye blinks and jaw clenches annotation data, dropped samples, etc.). Then, we subtract the 30 s baseline (relaxation phase of the study) from the rest of the signal, for each room and for each participant separately. Subsequently, we average the absolute band power of the five oscillations for each session in the VR as well as for each person, for later comparisons.

After the preprocessing of the absolute wave data and extraction of all 5 oscillation bands (using Formula 1), we conduct analysis of variance (ANOVA) with Sidak's correction for multiple comparisons, where the mean oscillation values for each location are the dependent variables, and the type of rooms and type of oscillation are independent variables.

$$alpha_{absolute} = (alpha_{TP9} + alpha_{AF9} + alpha_{AF9} + alpha_{TP10})/n - alpha_{basseline} \quad (1)$$

n = real active number of channels for each participant

The mean oscillation values (Formula 2), of each room assigns the emotion labels – calm (0) and active (1) – of the room for the virtual spaces and the sentences describing the rooms.

$$\bar{\alpha} = 1/25 \sum_{i=1}^{25} \alpha_i \quad (2)$$

The text sentences of each room transcribed from the audio are labeled with the room emotion data for further correlation analysis and NLP classification training. Basically, for each room (10 in total) in VR, 450 sentences from the subjects (26 in total) with 2 labels, calm and active from the EEGs. Correlation matrices are then be generated for all 5 oscillation bands and rating differences to uncover possible correlations between the emotions, sentences, and room parameters.

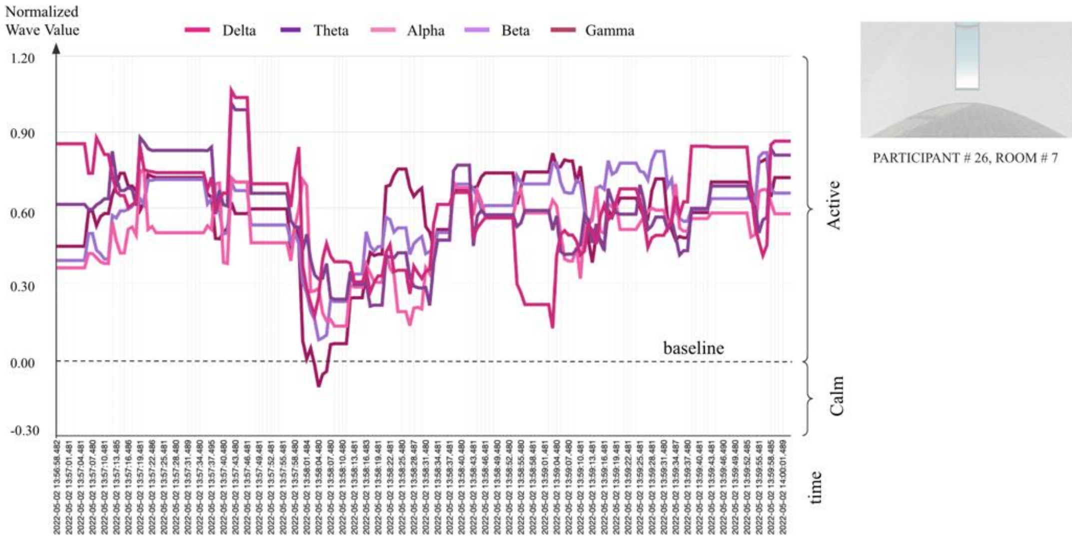


Fig.5: Example of emotion labeling

When participant experiencing the room number 7, we record his brain wave, and audio sentences stimulated by this room. We find that, he had an active emotion that above the 0 line – which is his baseline of brain wave. So, we label this sentence with active.

4.3. NLP Classification

Aside from quantitatively analyzing, our dataset can also be used to define natural language processing (NLP) tasks. Bidirectional Encoder Representations from Transformers (BERT) is a deep learning-based model for NLP which are pre-trained using a very large text corpus (Devlin et al., 2019). BERT is used for text classification training to compare the spatial emotions in virtual reality from our dataset and non-spatial emotions from the existing emotion dataset, such as SemEval2018Task1 (Mohammad et al., 2018), Ascertain (Subramanian et al., 2018) and Dreamer (Katsigiannis and Ramzan, 2017).

In this section, we introduce two novel tasks based on our dataset. In the first task, given a space's parameters of shape, length, width, and height, we predict whether the room will arouse a calm emotion or an active one. In the second task, given a sentence, we predict the emotions aroused by the spaces. Both these tasks are not only challenging from an NLP perspective, but also have potential applications. For example, models for predicting the emotions aroused by a space might be used in recommendation systems for architectural design. Also, a model trained to predict the emotions given comments using thousands of training examples might result in better design feedback for architects.

In task 1, since there are 10 rooms and 4 parameters, but there are only 1402 sentences, the mAP of the model is not high - only 0.3. For classification training with 10 labels, a larger training set should be required to learn the features.

In task 2, the model has an F1 score of 0.82 for the active label in the validation set while an F1 score of 0.37 for the calm label. This is since the number of labels in the overall training set and validation set is about 50% less than that of the active (Figure 3), which makes the samples challenging to learn and validate. Therefore, in the results of classification learning, the machine tends to classify more sentences as active.

The results on the two classification tasks show that the discrete distribution of the training set over room parameters and the imbalanced distribution over sentiment labels lead to biased training results. The envisioned solutions include increasing the number of participants in sample collection to increase the dataset and balancing the number of two types of sentiment labels.

Table 7: Results of trained BERT model of Task 2

	precision	recall	f1-score	support
Calm	0.35	0.41	0.37	24
Active	0.69	1.00	0.82	54
accuracy			0.69	78
macro avg	0.35	0.50	0.41	78
weighted avg	0.48	0.69	0.57	78

4.4. Dataset Comparison

We compare our dataset with existing sentiment classification datasets in two aspects: psychological datasets and NLP emotion datasets. Compared to other psychology datasets, our dataset is the first to use space as a medium for natural language and EEG. Existing datasets inspire emotions through images and quantify the data by measuring data to build datasets, such as Ascertain (Subramanian et al., 2018), Dreamer (Katsigiannis and Ramzan, 2017), and FER-2013. However, our dataset is a language corpus inspired by spaces and measured by EEG. Such datasets are difficult to analyze because language is ambiguous, and the mutual interpretation between emotion and language is also ambiguous. Also, the language of everyday life is not extreme and does not have a strong emotional consensus to provide further analysis. Compared with other NLP sentiment classification datasets, on the one hand, our dataset provides different sentiment labels since EEG devices are better at accurately measuring the emotional state of calm and active. If complex emotional states, such as happiness, need to be measured, more complex measurement equipment is often required. Therefore, our dataset adopts different emotion labels from existing NLP emotion classification datasets in the use of emotion models. For example, SemEval2018Task1 (Mohammad et al., 2018) adopts emotion labels such as anger, fear, joy, or sadness, while we use calm and active. On the other hand, our dataset comes from precise linguistic descriptions of spaces, and other datasets, such as SemEval2018Task1 (Mohammad et al., 2018), have sentences from Twitter – a language corpus that describes elements of every aspect of life. Since spatial elements inspire the sentences in our dataset, our dataset is a more spatially relevant corpus.

Table 8: Results of trained BERT model of Task 2

Sentences	Our dataset	SemEval-2018 Task 1: Affect in Tweets
This space feels more relaxing, but with more sounds like the space for worshipping or for prayers.	calm	joy
I feel very energetic, and I feel like dancing.	active	joy
I can't stop. I finished - dejected. luckily no one is in the bathroom. So, I go to a stall and wait until my pants are dry.	active	fear
Well stock finished & listed, living room moved around, new editing done & fitted in a visit to the in-laws. #productivityatitsfinest #happy	active	joy

5. DISCUSSION AND LIMITATIONS

In settings of VR spaces, on the one hand, the parameters are more controllable than real-world spaces, and VR spaces are a cheaper experimental test tool than real-world constructions. On the other hand, VR spaces may not necessarily reflect real-world user experiences. Spatial elements in the real world are complex and changeable, and selecting the combination of complex architectural features is key to experimental testing. For example, some researchers have selected elements such as openness and closeness (Ergan et al., 2018a). The more choices of elements, the more complex the spatial composition and the longer the experiment.

Participants in this study experienced the space at a fixed point standing or sitting. However, in the real world, the experience of space is continuous, and people experience architectural spaces not statically, but with movements. As a next step, moving panorama video may be a good start for testing real-world spaces.

The study tested brainwave data from 26 participants in a VR space using a Muse 2 device with 4 electrodes. More expensive EEG equipment may be able to reflect more comprehensive brain wave data. For example, the 14-channel EEG device (Emotiv Epic) with O2 and P7 electrodes (Shemesh et al., 2017), is a particularly important signal point (citation) for spatial experience. In addition, in terms of body sensors to capture more bio data, other devices can be used, such as electromyography (EMG), galvanic skin response (GSR), heart rate in photoplethysmogram (PPG), and eye tracking. For instance, the user's gaze trajectories or other behaviors in VR spaces may also cause measurement biases in the data for emotions.

During the experiment, this study collected participants' data such as EEG and audio, while the participants were sitting or standing and speaking or listening. Certain factors, such as the sitting or standing state of the participants, should be controlled. Participant 25, for example, said that speaking in the VR space would arouse his excitement. Other factors, such as smell and white noise, and psychological cognitive factors including memory, will also cause certain experimental errors in the impact of spatial emotions. In terms of NLP, we asked the participants particular questions for feedback on their emotional states. However, many kinds of questions might be asked and answered. Future research should include a spatial task with guidelines, such as asking participants to wait, to stimulate specific feelings (stress or calm).

6. CONCLUSION

Our spatial dataset, the first publicly available spatial dataset for spatial research and design purposes, contains 1402 sentences and 2 labels – calm and active. Our analysis shows interesting trends and predicts two results: (i) the emotional reaction to a space based on spatial features and (ii) the emotion in brain waves corresponding to a sentence based on its contents. Our experiments show that certain properties of spaces, such as length, width, and height, are correlated with higher possibilities of calmness, reflected in EEG data. We offer a useful NLP emotion classification dataset for architectural design improvement using everyday sentences. More importantly, we hope that our dataset helps architects understand the virtual spatial emotions in everyday language to guide design. In terms of application, our model can generate emotional evaluations of spaces based on the user's verbal description and brainwave data. Our models can be used to evaluate machine-learned or human-generated designs in the age of AI.

ACKNOWLEDGMENTS

This material and devices are based upon courses and work supported by the MIT Architecture, Representation and Computation group (ARC). The experimental results come from experiments' participants from graduate students and professors of MIT and Harvard, and the views, opinions, and/or findings expressed are those of the author(s).

REFERENCES

- BROWN, K.W., RYAN, R.M., 2003. The benefits of being present: Mindfulness and its role in psychological well-being. Journal of Personality and Social Psychology 84, 822–848. https://doi.org/10.1037/0022-3514.84.4.822

- DEVLIN, J., CHANG, M.-W., LEE, K., TOUTANOVA, K., 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv:1810.04805 [cs].
- DIEMER, J., ALPERS, G.W., PEPERKORN, H.M., SHIBAN, Y., MÜHLBERGER, A., 2015. The impact of perception and presence on emotional reactions: a review of research in virtual reality. Frontiers in Psychology 6.
- EL BEHEIRY, M., DOUTRELIGNE, S., CAPORAL, C., OSTERTAG, C., DAHAN, M., MASSON, J.-B., 2019. Virtual Reality: Beyond Visualization. Journal of Molecular Biology 431, 1315–1321. https://doi.org/10.1016/j.jmb.2019.01.033
- ERGAN, S., RADWAN, A., ZOU, Z., TSENG, H., HAN, X., 2018a. Quantifying Human Experience in Architectural Spaces with Integrated Virtual Reality and Body Sensor Networks. Journal of Computing in Civil Engineering 33. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000812.
- ERGAN, S., SHI, Z., YU, X., 2018b. Towards quantifying human experience in the built environment: A crowdsourcing-based experiment to identify influential architectural design features. Journal of Building Engineering 20, 51–59. https://doi.org/10.1016/j.jobe.2018.07.004
- HU, X., CHEN, J., WANG, F., ZHANG, D., 2019. Ten challenges for EEG-based affective computing. Brain Science Advances 5, 1–20. https://doi.org/10.1177/2096595819896200
- KATSIGIANNIS, S., RAMZAN, N., 2017. DREAMER: A Database for Emotion Recognition through EEG and ECG Signals from Wireless Low-cost Off-the-Shelf Devices. https://doi.org/10.1109/JBHI.2017.2688239
- MOHAMMAD, S., BRAVO-MARQUEZ, F., SALAMEH, M., KIRITCHENKO, S., 2018. SemEval-2018 Task 1: Affect in Tweets, in: Proceedings of the 12th International Workshop on Semantic Evaluation. Presented at the Proceedings of the 12th International Workshop on Semantic Evaluation, Association for Computational Linguistics, New Orleans, Louisiana, pp. 1–17. https://doi.org/10.18653/v1/S18-1001
- SHEMESH, A., LEISMAN, G., BAR, M., GROBMAN, Y.J., 2021. A neurocognitive study of the emotional impact of geometrical criteria of architectural space. Architectural Science Review 64, 394–407. https://doi.org/10.1080/00038628.2021.1940827
- SHEMESH, A., TALMON, R., KARP, O., AMIR, I., BAR, M., GROBMAN, Y.J., 2017. Affective response to architecture – investigating human reaction to spaces with different geometry. Architectural Science Review 60, 116–125. https://doi.org/10.1080/00038628.2016.1266597
- SUBRAMANIAN, R., WACHE, J., ABADI, M.K., VIERIU, R.L., WINKLER, S., SEBE, N., 2018. ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors. IEEE Transactions on Affective Computing 9, 147–160. https://doi.org/10.1109/TAFFC.2016.2625250
- SUHAIMI, N.S., MOUNTSTEPHENS, J., TEO, J., 2020. EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities. Computational Intelligence and Neuroscience 2020, e8875426. https://doi.org/10.1155/2020/8875426