Descriptive Linguistic Patterns of Group Conversations in VR

Cyan DeVeaux* Stanford University, USA
David M. Markowitz† University of Oregon, USA
Eugy Han‡ Stanford University, USA
Mark Roman Miller§ Stanford University, USA
Jeffrey T. Hancock¶ Stanford University, USA
Jeremy N. Bailenson‖ Stanford University, USA

ABSTRACT

Although talking is one of the most common activities in social VR, there is little empirical work identifying what people say and how they communicate in immersive, virtual settings. The current paper addresses this opportunity by performing automated text analysis on over 4,800 minutes of in-VR, small group conversations. These conversations took place over the span of two months during a university course where 171 students attended discussion sections via head-mounted displays. We provide a methodology for analyzing verbal communication along two dimensions: content and structure. We implement methods to describe linguistic patterns from the class and introduce a preliminary VR Dictionary.

Keywords: virtual reality, social virtual reality, language

Index Terms: Human-centered computing—Collaborative and social computing—Empirical studies in collaborative and social computing

1 INTRODUCTION

Social virtual reality (VR) is an emerging ecology of platforms where users are represented by avatars who are networked into a variety of scenes and activities. Users experience a novel form of avatar-mediated interaction where physical body movements are tracked and rendered via avatars while immersed in a virtual environment. As a result of social VR's unique positioning between face-to-face and digital communication, there is relevance to understanding language patterns through and within VR. Pfieffer (2012) outlines how VR can be leveraged to effectively test theories of multimodal interaction in linguistics. It enables a customized simulation of communication scenarios by modifying stimuli in virtual environments [4].

However, there has been little empirical work investigating verbal communication in immersive, naturalistic contexts over time. The aim of the current study is to understand natural language use in social VR through performing automated, exploratory text analysis of over 4,800 minutes of group conversation taken over the course of seven weeks during a university course. We describe patterns in the content and structure of language that took place over time to examine the dominant content themes discussed and how students shared speaking time in a social VR classroom.

2 METHODS

Participants were 171 university students enrolled in a 10-week course about VR. At the beginning of the course, students were invited to participate in an IRB-approved study of how repeated exposure to VR influenced their individual and group behavior. Participant consent went through a rigorous process, approved by two separate organizations within our university. Students had an interactive, hour-long discussion of the study procedures and data collection before deciding to consent. The IRB process required that researchers and course staff did not know which students opted-in as participants until after the course so that there would be no plausible appearance of coercion to participate in the study. Therefore all students were recorded in this study, but only data associated with consenting participants was used (n=158). Participants were reminded of the recording through a visual notification at the start of a recorded session or upon joining a recorded session. A 3rd party arbitrar oversaw data collection during the course. Each participant was provided with a Meta Quest 2 headset and attended weekly 30-minute discussion sections on ENGAGE, a collaborative social VR platform. A total of 24 groups, 5-8 members each (M = 6.71, SD = 0.81), met weekly for eight weeks and were led by one of three instructors. Each section, with the exception of Week 5 which was removed from the scope of this data, had a similar format consisting of full-group discussion, an individual creativity activity, and a show-and-tell of their creation. Audio in sections was non-spatialized.

Of the 168 sections that occurred across eight weeks, 162 sections were recorded using ENGAGE's recording feature, saved as .myrec files, converted into audio files, and transcribed using an automated text transcription software, Otter.ai. The remaining six sections were unusable due to technical errors. Research personnel manually edited the transcriptions for accuracy and to redact speech from non-consenting participants. Words from transcripts were quantified with Linguistic Inquiry and Word Count (LIWC) [3], an automated text analysis software. We applied the Meaning Extraction Method [1] with this tool and used a Principal Component Analysis (PCA) with varimax rotation to evaluate the content words that clustered statistically to form themes. Our analysis retained unigrams, bigrams, and trigrams that appeared in at least 10% of texts and loadings $>|.20|$. The unit of analysis was language that occurred within each section of each week. Only student language was retained in these analyses and we therefore excluded members of the teaching team.

To obtain speaking time data of students, we extracted how loudly a participant was speaking in a given time frame using a floating-point value from zero to one. We selected a value of 0.001 as a threshold for speaking. The total number of frames above this threshold for each participant indicates speaking time. Speaking time of students varied from 0 minutes to 37.11 minutes (M = 2.84, SD = 4.07). We excluded sections that included individuals with unusually high speaking time from this analysis. That is, considering the structure of sections and that recordings were typically 30 to 40 minutes in length, we set 20 minutes as the cut-off.

Using this process, we measured the *linguistic content* and *linguistic structure* of over 164,000 words. Linguistic content used the PCA data to identify what people spoke about in VR. Linguistic structure used the entropy of speaking time distributions in each section to measure speaker dominance or evenness.

*e-mail: cyanjd@stanford.edu
†e-mail: dmark@uoregon.edu
‡e-mail: eugyoung@stanford.edu
§e-mail: mrmillr@stanford.edu
¶e-mail: hancockj@stanford.edu
‖e-mail: bailenso@stanford.edu

C1: Avatar Embodiment		C2: Sensory Processes		C3: Class Artifacts		C4: Learning	
$\lambda = 7.451$	% = 5.961	$\lambda = 5.075$	% = 4.060	$\lambda = 3.889$	% = 3.111	$\lambda = 3.759$	% = 3.007
Word	Loading	Word	Loading	Word	Loading	Word	Loading
avatars	0.704	room	0.590	name	0.608	space	0.525
hands	0.668	sitting	0.497	class	0.569	understand	0.498
avatar	0.665	feels	0.492	learn	0.547	helpful	0.477
realistic	0.619	nice	0.477	technology	0.459	learning	0.463
body	0.476	zoom	0.414	zoom	0.427	better	0.459
look	0.461	sounds	0.403	guys	0.421	interactive	0.439
weird	0.440	presence	0.391	people	0.404	hard	0.393
C5: Future		C6: Real Life		C7: Gaming		C8: Cool Factor	
$\lambda = 3.240$	% = 2.592	$\lambda = 3.128$	% = 2.502	$\lambda = 2.981$	% = 2.384	$\lambda = 2.818$	% = 2.254
Word	Loading	Word	Loading	Word	Loading	Word	Loading
change	0.616	real life	0.710	games	0.643	said	0.511
effects	0.518	life	0.689	game	0.638	great	0.507
important	0.392	real	0.674	playing	0.533	pretty cool	0.505
future	0.356	place	0.277	play	0.514	thought cool	0.385
show	0.353	sense	0.275	necessarily	0.313	quick	0.344
video	0.343	bit	0.267	video	0.294	cool	0.322
thinking	0.319	find	0.265	type	0.293	trying	0.316

Table 1: *Extracted Components from the Meaning Extraction Method*

3 Results

Linguistic Content. We used scree plot evidence, variance explained, and thematic interpretability in deciding the number of themes to extract. In total, eight components were retained (Table 1): avatar embodiment, sensory processes, class artifacts, learning, future, real life, gaming, and cool factor. Some themes may be specific to an immersive VR classroom, such as information about the class (C3). However, other themes may be domain-independent such as "avatar embodiment", "learning", and "cool factor", where such themes might be representative of what people talk most about while in virtual spaces.

Linguistic Structure. As Figure 1 demonstrates, the distribution of speaking times differs drastically across groups. On the right side of the figure, the amount of speaking time was typically centered around one or two speakers, while on the left side, there is greater entropy within the conversation. In our study, we had at least 25 discussion sections where a singular student accounted for more than half of the total student speaking time. While conversational dominance has been found in desktop-based virtual worlds that leverage voice chat [2], we extend these findings to immersive virtual environments.

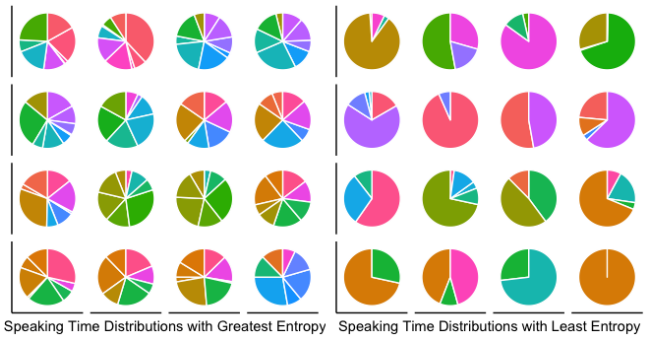

Figure 1: *Speaking Distributions by Greatest and Least Entropy*

4 Discussion & Conclusion

In this paper, we observed what students communicated about and how much they communicated in a class in VR about VR. There were eight reliable themes that students spoke about in VR across weeks (See Table 1). These themes and their corresponding language represent a preliminary VR dictionary, a tool that can be used to understand and measure what people talk about when immersed in VR. For example, avatar embodiment topic was the most robust theme in terms of variance explained, demonstrating that students in a virtual class may tend to linguistically focus on virtual representations of self. It will be critical for future work to identify how focus on one's avatar might also be associated with other social and psychological dimensions. Our initial VR dictionary provides an opportunity to understand how people are thinking, feeling, and what they are focusing on psychologically in VR. We also found varying distributions in speaking time, with a number of them reflecting conversational dominance by one person. This reflects dynamics that can also occur in offline spaces, where some people tend to control conversations more than others. Therefore, people may communicate and treat conversation in VR like they do in the physical world [5]. Through examining linguistic content and structure, the current work demonstrates how words are used in social VR.

The study has several limitations. The preliminary categories identified for the dictionary were influenced by the VR-centric topics from the course that the conversations took place in. Also, discussion sections were mainly facilitated by course instructors and therefore did not involve much casual conversations. Future research should examine conversations that occur outside of the context of a course and across different communication media. Future work should also examine how psychological cues in language relate to self-report measures.

References

[1] D. M. Markowitz. The meaning extraction method: An approach to evaluate content patterns from large-scale language data. *Frontiers in Communication*, 6:13–13, Feb. 2021. Publisher: Frontiers.

[2] L. Newon. Multimodal creativity and identities of expertise in the digital ecology of a world of warcraft guild. *Digital discourse: Language in the new media*, 131, 2011.

[3] J. W. Pennebaker, R. L. Boyd, R. J. Booth, A. Ashokkumar, and M. E. Francis. Linguistic Inquiry and Word Count: LIWC-22, 2022.

[4] T. Pfeiffer. Using virtual reality technology in linguistic research. In *2012 IEEE Virtual Reality Workshops (VRW)*, pp. 83–84, Mar. 2012.

[5] B. Reeves and C. Nass. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press, United Kingdom, 1996.