TEACHING with TECHNOLOGY 2020

Selected papers from the JALTCALL2020 Conference

JALTCALL 2020

Edo Forsythe, Editor-in-Chief

Associate Editors: Anthony Brian Gallagher
Douglas Jarrell
Branden Kirchmeyer
Louise Ohashi
James York

ISBN 978-4-901352-63-5

Teaching with Technology 2020
Selected papers from the JALTCALL2020 Conference

JALT CALL is a Special Interest Group (SIG) of the Japan Association for Language Teaching (JALT) that focuses on Computer-Assisted Language Learning (CALL) and technology in language learning.

http://jaltcall.org/

JALT is dedicated to the improvement of language teaching and learning.

JALT Central Office
Urban Edge Bldg. 5F, 1-37-9 Taito,
Taito-ku, Tokyo 110-0016, Japan

http://jalt.org/

First published in Japan in 2021 by the JALT CALL SIG.

Layout by Paul Mason

https://doi.org/10.37546/JALTSIG.CALL2020.9

9

The Effect of Modality on Oral Task Performance in Voice, Video, and VR-based Environments

Hayato Tokutake, James York, and Hiroshi Nakayama, Tokyo Denki University

Abstract

Synchronous computer-mediated communication (SCMC) is a topic of great interest in CALL literature where research has investigated the effectiveness of SCMC compared to traditional face-to-face instruction. However, there are few studies that investigate the intrinsic differences in SCMC modes, particular in terms of their effect on oral communication. At the JALTCALL 2019 conference, we introduced research which assessed the anxiety-reducing affordances of VR. This year we presented results of a follow-up study which focused on the effect of SCMC modality on learners' speaking performance.

30 participants (15 pairs) completed a spot-the-difference task within three different SCMC modes: voice, video, and virtual reality (VR). Using the complexity, accuracy, and fluency (CAF) model, participants' oral task performances were analysed. Results suggest that the voice mode promoted the highest structural complexity, however, the VR mode promoted the highest lexical complexity. Findings therefore suggest that different modes of communication may be used to focus on different skill development. Additionally, practitioners should consider how modality affects learner anxiety and choose the most appropriate system for their students and needs. This paper introduces the VR system, a detailed analysis of results, pedagogical implications, and future research directions for the use of VR in language teaching contexts.

コンピュータを介した同期型コミュニケーション（SCMC）は、CALLの文献で大きな関心が寄せられているトピックであり、従来の対面式の指導と比較してSCMCの有効性が調査されている。しかし、特にオーラルコミュニケーションへの影響の点で、異なるSCMC間での違いを調査する研究はほとんどされていない。JALTCALL 2019では、VRにおける不安軽減アフォーダンスを評価する研究を紹介した。今年は、学習者のスピーキングパフォーマンスに対するSCMCのモダリティの効果に焦点を当てたフォローアップ調査の結果を示した。

本研究では、30名の被験者（15ペア）に、Voice、Video、Virtual Reality（VR）の3つの異なるSCMCで間違い探しのタスクを行わせた。本研究では、複雑さ、正確さ、流暢さ（CAF）モデルを使用して、参加者のスピーキングパフォーマンスを分析した。実験の結果、文法の構造的複雑さにおいて、Voiceモードで最もスコアが良くなった。一方で字句の複雑さにおいては、VRモードが最もスコアが良くなった。

したがって、スキルの開発に焦点を当てる場所によって、異なる通信モードそれぞれに可能性がある

ことが示唆された。そのため、教育者はSCMCのモダリティが学習者の不安にどのように影響するかを検討し、学生の状況に合わせて最適なシステムを選択する必要がある。

この論文では、VRシステム、結果の詳細な分析、教育的影響、および言語教育のコンテキストでVRを使用するための将来の研究の方向性を紹介する。

Keywords: **SCMC, VR, modality, oral proficiency, CAF**

There have always been critiques aimed at the normalization of English as a lingua franca (see Pennycook, 2000), and, entering the third decade of the 21st Century, this issue is as pertinent as ever (Su, 2016). However, the fact remains that speakers of different nationalities may often default to English when communicating with each other. As such, in an increasingly globalized and English-speaking world, the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT) has been promoting the development of practical English skills since 2014. At the primary, secondary, and tertiary levels, there is an emphasis on students being able to "assertively make use of their English skills, think independently, and express themselves" (MEXT, 2014, p. 3).

The spread of English as a Lingua Franca, and intercultural communication in general, may be linked to the rapid development of our networked society and the technology that underpins the vast majority of our daily, increasingly computer-mediated communication. Subsequently, it is not surprising that within foreign and second language studies, there has been a rapid increase in both research and practical implementation of computer-mediated communication (CMC) (Lin, 2014).

This study focuses on the affordances of three different synchronous CMC (SCMC) modalities and their effect on learners' oral task performance. In particular, the study aims to understand what the specific affordances of virtual reality (VR) technology have on learner output as they complete a dialogic task within the VR environment. In order to ascertain whether VR promotes a unique set of communicative affordances, student output is compared with two more frequently studied modalities: voice and video SCMC.

Literature Review

SCMC Studies

CALL research exploring the potential of CMC for language learning has progressed in accordance with technological advances. An early study by Warschauer (1996) explored how telecommunication within a synchronous, text-based CMC domain affected learner motivation and willingness to communicate in comparison to face-to-face (FTF) communication. Findings suggested that students participated in conversations more equally in the CMC mode with an increase in lexical complexity and formality. This study therefore provided empirical evidence that modality does indeed affect learner output. As a critique of the study, and related to the present study, it is not surprising that there was a difference in lexical complexity for both modes, as this difference exists between written and oral output before even considering CMC.

At the time of Warschauer's study, synchronous CMC modalities that incorporated voice

and video were not in widespread use. However, since then, such studies (i.e., those that compared voice and video-based CMC with FTF communication) have started to appear. Voice and video modes are more similar to FTF in that they primarily rely on oral communication. The word *primarily* is used in the previous sentence because the majority of voice and video-based CMC tools emerged as an extension of preexisting text-based modes. For example, although Skype is a VOIP-based voice and video communication tool primarily, it is multimodal in that it also supports text-based interactions. One particular study which acted as the impetus for a previous study of our own (York et al., 2021) was that of Satar and Özdener (2008). They investigated the effects of SCMC on speaking proficiency and anxiety. Two modes were employed: text and voice SCMC. Sixty participants completed language activities in one of the two modes in addition to their usual classroom-based instruction. A control group of 30 students did not complete any of the additional activities, thus a total of 90 participants were used in the study. Based on speaking proficiency tests at the end of the treatment, it was discovered that the speaking proficiency of both of the experimental groups increased in comparison to the control group in a post-experiment speaking test. Thus, one important finding of this study was that the text-based SCMC group were able to improve their speaking proficiency without physically practising speaking. However, in comparison with the present study, Satar and Özdener's study lacked a rigorous evaluation of learner output *during* task performance, only their ability to perform during a post-task test.

Following on from Satar and Özdener's study, Yanguas (2010) added video-based SCMC as a modality in a study which explored how the addition of video affected learner output. Fifteen learner dyads (i.e. pairs) were created and assigned to complete an oral communication task within a voice-based, video-based, or face-to-face modality. The study assessed learner output performance in terms of the number of times they negotiated for meaning (NfM) with their partner. Findings suggested that there was no difference in terms of NfM between the video-based SCMC mode and the FTF mode, however, there was an increase in NfM for the dyads that completed the task in the voice-based SCMC mode. The reason for this difference was attributed to the lack of visual stimulus, meaning that the dyads had to rely on verbal explanations instead.

In addition to the tools outlined above (voice, video, and text based SCMC), some studies have investigated the affordances of SCMC within virtual worlds (VWs), and digital games, which are becoming increasingly sophisticated (Melchor-Couto, 2017; 2018; Wigham & Chanier, 2015). York (2019) operationalized tasks for one such virtual world: *Minecraft*. In a counterbalanced, repeated-measures study, 10 dyads carried out three sets of oral communication tasks in two different environments: within *Minecraft* and FTF (totalling 6 tasks). Participants' oral proficiency was analysed in terms of output complexity, accuracy and fluency, and it was found that the VW hindered fluent output, but that modality had no significant effect on accuracy and complexity dimensions. Instead, task *type* appeared to be a more influential factor. Results echoed a similar study by Yanguas and Bergin (2018), who found no difference in the number of learning related episodes (LREs) when participants completed jigsaw tasks in video and voice based SCMC modes.

Finally, related to the use of virtual worlds as domains for CMC, there are a number of meta-analyses on the use of SCMC in language learning and teaching environments.

Focusing on one recent analysis, Zeigler (2016) analysed the results of 14 studies which compared the effectiveness of SCMC to that of FTF communication. Studies were selected based on their inclusion of interaction between participants during task performance. Findings of the meta-analysis found that overall, there was no significant effect between SCMC and FTF on learner interaction. Notably, of the 14 studies chosen, none featured the use of a virtual domain and only three implemented oral SCMC as the modality of focus. This highlights a lack of focus and a need for more studies which operationalize tasks within an oral SCMC mode. The present study hopes to fill this gap by utilizing three different SCMC oral modes.

VR and Second Language Interaction

VR and its application in language learning is a nascent field, with few empirical studies to date. The majority of research papers on the use of VR are hypothetical and based on the possible affordances of the technology in language learning and teaching contexts. Hawkinson et al. (2017) introduced several different reality-augmenting technologies, differentiating between augmented reality, virtual reality, and mixed reality modes. Following, they connected these technologies to cognitive science studies, introducing the potential of such tools for language learning. Similarly, Bonner and Reinders (2018) introduced a number of ways that VR could be utilized in classroom teaching, which is a first step to its implementation. Their ideas included virtual campus or city tours, as well as practicing public speaking via specific apps.

Regarding empirical studies, York et al. (2021) researched the affective affordances of VR as a domain for second language interaction. Their study assessed the effect of three different SCMC modes on learners' foreign language anxiety (FLA): voice, video, and VR. Results of quantitative data analysis revealed that all three modes reduced learners FLA compared to pre-test FLA scores, however, learner perceptions suggested that of the three environments VR was considered the most effective domain for language learning, as well as being the easiest to communicate with interlocutors. Findings therefore align somewhat with those of Melchor-Couto (2017) who found that learner FLA was reduced when learning within the virtual world Second Life compared to an FTF group. The reduction in FLA when communicating in virtual domains has been attributed to the *avatar effect* where the avatar provides users with an anonymous shield through which to talk (Lotherington & Jenson, 2011). Subsequently, Xie et al. (2019) explored the affordances of a VR system on oral proficiency. Two groups, one with the aid of a VR environment and one without, performed a monologic, oral task. Results suggested that both the content and vocabulary of participants' oral were influenced by mode. The VR group had statistically significantly higher scores than the group without the aid of the VR environment, and the researchers concluded that access to additional stimuli may have influenced the higher scores. In summary, research on VR has typically proposed practical ways in which it could be implemented in classroom contexts, though there are few empirical studies appearing in the literature.

Research Question

The aim of the current research is to uncover the effect of SCMC modes on oral task performance. As such, the research question is:

1. Is there a difference in learner oral proficiency when completing interactive tasks within voice, video and VR-based SCMC modes?

As a follow up, sub-question: If there is a difference in performance based on modality, what are the affordances for each domain for promoting language proficiency?

Methodology

Participants

The study was undertaken at a private science and engineering university in Saitama, Japan. Thirty participants volunteered to take part in the study. Their mean age was 20.9 (*SD* = 1.12). They were all native speakers of Japanese and had received the same amount of formal English education, thus making this a homogeneous group of learners (see Table 1). All participants provided their informed consent via a consent form at the start of the study. The form outlined the study purpose and data handling procedures including information regarding the anonymization of data.

Table 1

Baseline Characteristics of Participants

Baseline Characteristic		n	%
Age			
	19	2	6.7
	20	11	36.7
	21	9	30.0
	22	6	20.0
	23	1	3.3
	24	1	3.3
Gender			
	Male	26	86.7
	Female	4	13.3
Native Language			
	Japanese	30	100

In order to avoid a familiarity effect, the researchers created dyads by pairing participants with partners that they did not know and met for the first time on the day of the study.

Dyads completed a spot-the-difference task in all three domains, and so a counterbalance design was used to prevent the resultative effect in which task familiarity increases a student's performance and confidence in their performance (see Skehan, 1991).

Additionally, dyads completed the tasks in different orders in a counterbalanced, repeated-measures design. Three groups were created, and are referred to as Group A, B, and C.

Instruments

The three SCMC modes utilized in this study were oral (henceforth voice), video, and VR (developed in Unity **https://unity.com/**) A spot-the-difference task was created in the VR domain and the same assets were reused in both the voice and video and SCMC modes in order to unify the content for each task.

VR modality

Figure 1

An Example of the Spot-the-Difference Task in VR (Participant 1)

Figure 2

An Example of the Spot-the-Difference Task in VR (Participant 2)

As can be seen in Figures 1 and 2, the spot-the-difference task featured a doll's house that each participant could look into, as well as seeing their interlocutor (Figure 3). The number of differences in each task was operationalized to three per task. Avatars were utilized as embodied versions of the participants. These avatars echoed participants' body, head, and arm movement. This was achieved with the use of the HTC Vive head mounted display unity (see Figure 4).

Figure 3

A Screenshot of the VR Task Showing the Interlocutor

Figure 4

A Participant Wearing the HMD During the VR Task

Voice and Video Modalities

For the voice and video modalities, the position of items within the VR doll houses were changed and screenshots taken to create different versions of the spot-the-different task.

These screenshots were presented to learners on a monitor screen (see Figure 5). The difference between the voice and video modalities was whether interlocutors were visible or not (see Figure 5 and Figure 6).

Figure 5

A Participant Completing the Spot-the-Difference Task in the Voice Modality

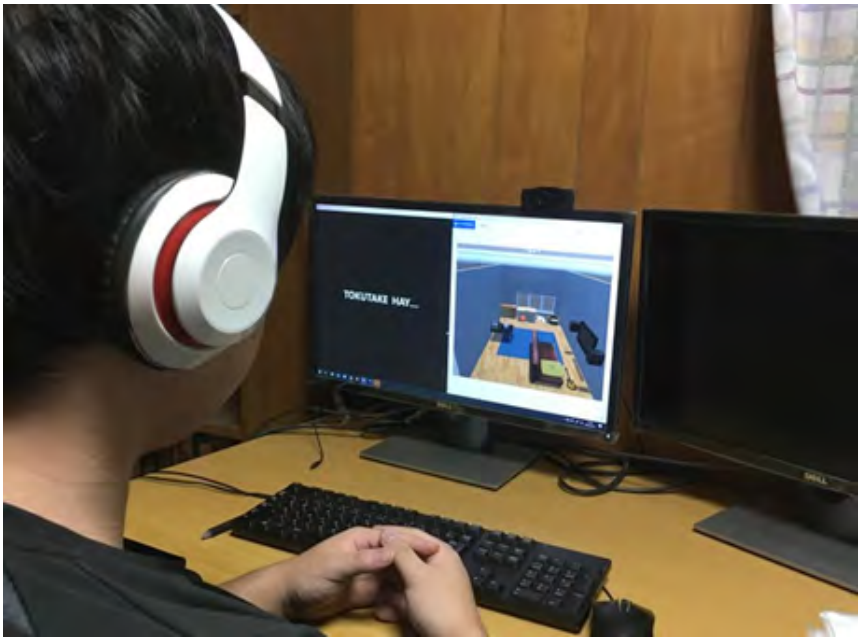


Figure 6

A Participant Completing the Spot-the-Difference Task in the Video Modality

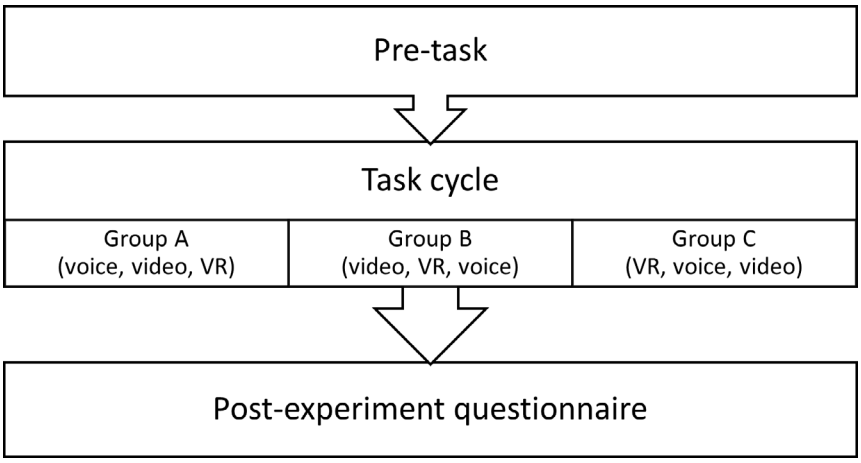
Zoom was employed as the communication software for all tasks, allowing for the recording of both audio and video.

Procedure

The experiment was a repeated measures design, where each pair of participants completed the spot-the-difference task in all three modes. The design is therefore based on that of York et al., (2020). Three groups of pairs (10 pairs per group) were created which completed tasks in a different order to avoid the Resultative Hypothesis (Skehan, 1991). That is, the mastery of tasks can affect learners' motivational attitudes and performance, and so the modality of task repetitions varied between groups (see Figure 7).

Figure 7

Experiment Flowchart



Before undertaking the spot-the-different tasks in each of the three SCMC modes, all participants completed a pre-task to prime them for the main task (see Appendix A). Upon completion of the pre-task, pairs carried out the spot-the-difference tasks in the order assigned to their group. The task cycle included an explanation of the task and instructions on how to complete the task within that domain (for instance, the VR task included a short demonstration on how to use the controllers to gesture to their partner).

Upon completing all three tasks participants answered questions on a post-experiment questionnaire. The questionnaire was devised to measure foreign language anxiety and collect perceptions of learning within each modality. For this paper, the only question that is referenced is the final open-ended question which allowed participants to comment on their experiences. This data was referenced as participant perceptions and used to provide further insight into any findings.

Data Analysis: The CAF Model

The CAF model is a robust way of assessing the proficiency of learner output along multiple, quantitative dimensions (Housen & Kuiken, 2009). The model examines learner output in terms of complexity, accuracy, and fluency (henceforth, CAF). The genesis of the model is associated with the work of Skehan (1998). There is a plethora of measures available in the literature for assessing learner output and so the following three sections will introduce those used in this study, including a rationale for their inclusion.

Complexity Measures

Learner output complexity can be measured in two ways: Structural complexity and lexical complexity. In this study, two methods for assessing output complexity were utilized, one for each complexity type.

- **Structural complexity** was measured by counting the number of words per utterance (Ortega, 1999).
- **Lexical complexity** was measured by the total number of different words a participant produced (number of types, not tokens).

The website Lextutor (**https://www.lextutor.ca/**) was utilized in accessing the number of types.

Accuracy Measure

Accuracy was measured by assessing the number of error-free utterances a participant produced. This measure resembles the most common measure found in the literature: error-free clauses (Skehan & Foster, 1997; Yuan & Ellis, 2003; Bygate, 2001). However, it was modified here due to the nature of participant speech. Participants did not produce utterances that contained multiple clauses (see also York, 2019).

Fluency Measures

There are two major measures for fluency: **temporal** and **vocal** fluency. Temporal fluency refers to a participant's rate of speech and is typically measured as words or syllables per minute (Yuan & Ellis, 2003). Vocal fluency refers to how "exact" a participant speaks, thus focusing on pauses, repetitions, and number of false starts (Skehan, 2009). In this study, participant fluency was measured as the number of words spoken per minute.

Statistical Tests

To prepare data for statistical tests, pairs were recorded and later the audio transcribed by the authors. The transcriptions were then coded for accuracy. Codes used were *Y* for an error-free utterance and *N* for an erroneous utterance. A sample can be seen in Table 2.

Table 2

Sample Transcription

Utterance Number	Participant Number	Utterance	Accuracy Codes
47	1	It is a difference point.	y
48	2	Yes	y
49	1	The guitar in front of sofa	n
50	2	Yes	y
51	1	OK	y
52	2	Where is the chest?	y
53	1	Chest is the desk left side.	n
54	2	Yes.	y
55	1	The same.	y
56	2	The same.	y
57	1	What color is carpet?	n

One-way repeated measures ANOVA analyses were utilized to look for statistically significant differences in mean scores for all measures. SPSS (version 24) was used to conduct analyses. Subsequently, a post hoc Bonferroni adjustment was used to understand which modalities' mean scores differed significantly. The alpha level for all statistical tests was set at $p < .05$.

Results

First, a detailed examination of participants' output in terms of complexity, accuracy, and fluency is presented. Based on any findings, the affordances for each mode are then explored.

Complexity

Mean scores for the number of **words per utterance** measure are displayed in Table 3. The highest mean score was recorded for the voice modality (2.82), the lowest was recorded for the VR modality (2.56).

Table 3

Mean Scores for the Number of Words per Utterance Measure

Modality	Mean	*SD*
Voice	2.82	0.473
Video	2.68	0.55
VR	2.56	0.398

Inspection of the results of the one-way repeated measures ANOVA revealed that there was a statistically significant interaction ($p < .05$), and pairwise comparisons revealed that there was a statistically significant difference between mean scores for the Voice and VR modes (Table 4 and Figure 8).

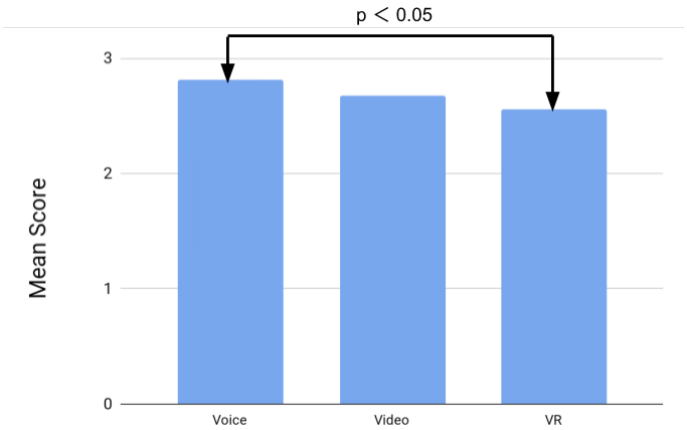
Table 4

Pairwise Comparisons for Simple Main Effects of Modality on Words per Utterance

Modality (I)	Modality (J)	Mean difference(I-J)	Sig
Voice	Video	0.14	0.23
Voice	VR	0.26	0.01*
Video	VR	0.12	0.18

Figure 8

Graphical Representation of Mean Scores for Number of Words per Utterance Measure



Regarding the second complexity measure: **number of different words spoken**, mean scores for each modality are presented in Table 5. The highest mean score was recorded for the VR mode, and the lowest for the video mode.

Table 5

Mean Scores for the Number of Different Words Measure

Modality	Mean Score	*SD*
Voice	31.23	9.62
Video	28.20	11.17
VR	36.97	17.25

Inspection of the results of the one-way repeated measures ANOVA revealed that there was a statistically significant interaction ($p < .01$), and so pairwise comparisons were conducted. Results showed a statistically significant difference between the video and VR modes (Table 6 and Figure 9).

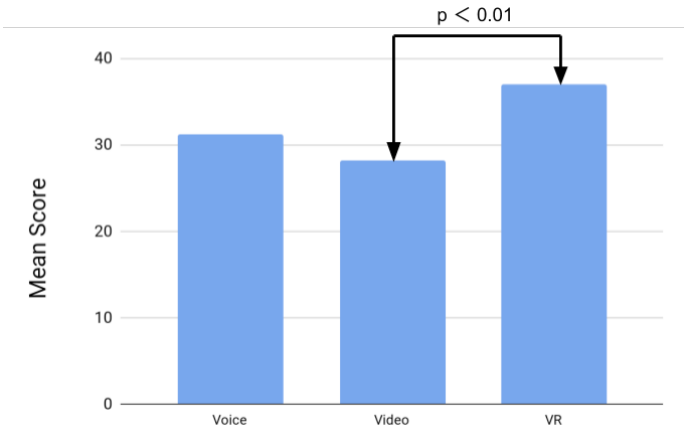
Table 6

Pairwise Comparisons for Simple Main Effects of Modality on the Number of Different Words Produced

Modality (I)	Modality (J)	Mean difference(I-J)	Sig
Voice	Video	3.03	0.36
Voice	VR	5.74	0.09
Video	VR	8.77	0.00**

Figure 9

Graphical Representation of Mean Scores for Different Words Measure



Accuracy

Table 7 shows the mean scores for participants error free utterances for each modality. Participants produced the most error-free utterances when completing the spot-the-difference task in the video mode, though all mean scores are similar. Inspection of the results of the one-way repeated measures ANOVA revealed no statistically significant interaction between mean scores.

Table 7

Mean Scores for Error-free Utterances Measure

Modality	Mean Scores	*SD*
Voice Chat	0.64	0.215
Video Chat	0.70	0.165
VR Chat	0.67	0.184

Fluency

Mean scores for the fluency measure employed in this study: **number of words per minute** are presented in Table 8. Though scores for each mode are similar, the highest mean

score was recorded for the voice mode, and the lowest for the VR mode. Inspection of the results of the one-way repeated measures ANOVA revealed no statistically significant interaction.

Table 8

Mean Scores for Participants' Number of Words per Minute

Modality	Mean Score	*SD*
Voice Chat	29.46	15.41
Video Chat	28.75	15.413
VR Chat	28.05	8.11

Post Experiment Questionnaires

Of the 30 participants, 26 gave comments to the open-ended question asking for feedback on their experiences. There were 18 comments in relation to the VR mode, 7 for video, and 2 comments recorded for the oral mode. A selection of typical comments for each mode are provided below.

A number of comments in relation to the VR mode mentioned that it was an effective domain for language learning, it was interesting, and the embodied experience was a catalyst for easier communication than other modes (see Table 9 for representative comments).

Table 9

Positive Comments Related to the VR Mode

Participant Number	Comment (translated from Japanese to English)
1	This was my first-time using **VR**, and I found it incredibly interesting.
2	I was able to see my partner's gestures when using the **VR** system, and so it felt like I was talking to somebody in front of me. It made me feel relaxed.
6	Speaking to someone in English using **video** or **VR** was a new experience for me and so I completed the tasks with a feeling of intrigue.
14	I was surprised that I didn't feel seasickness when completing the **VR** task.
16	**VR** was the most fun. I'm looking forward to a practical implementation of this system.
20	It was easier to complete the tasks with the **video** and **VR** systems because we could see our partner. With **voice** only, I didn't know when my partner was ready.
23	When using **video** the task screen and partner screen were separate, so it was easier to communicate in the **VR** domain because it was just one context.
29	I think **VR** will be very appealing to students because it is a currently hyped media. For that reason, I think it is an effective learning environment.

However, several negative comments were recorded for the VR domain (Table 10). These were related to the cognitive demands placed on learners due to the novelty of the

environment and the ease of gesturing in the domain meaning that they did not need to rely on their language skills. One comment mentioned that the VR mode placed increased spatial reasoning demands on participants, making it a more difficult domain to complete the task (Participant 9).

Table 10

Negative Comments Related to the VR Mode

Participant Number	Comment (translated from Japanese to English)
5	I was a little puzzled when using **VR** because I wasn't used to it.
9	It was hard to understand my partner's directions in **VR** because of the more complex spatial properties of the system
23	I was not very familiar with **VR** and so I didn't use many gestures. It was easier to gesture when doing the **video** task.
24	Because my partner was not directly in front of me during the **VR** task but off to the side, it was a little difficult to look over and check their gesture.
28	Because it was easier to gesture in **VR**, we didn't have to rely on our English ability so much..

Regarding the video mode, three comments stood out as negative opinions regarding this environment (Table 11). Comments generally mentioned that participants were more focused on their own image or doll's house and that they did not make use of the affordances of the video of their partner. Additionally, one comment mentioned feeling embarrassed by being visible to their partner.

Table 11

Comments Regarding the Video Mode

Participant Number	Comment (translated from Japanese to English)
12	Regarding the **video** and **voice** systems, although you could see your partner when using video, I spent more time looking at the picture, so I didn't look at my partner's face very much. I felt no difference between these two systems.
13	Speaking honestly, I didn't notice too much difference between the **VR** and **video** environments because I focused on my room rather than my partner.
23	During the **video** task, I couldn't look at my partner. I was also a little embarrassed to have my own **video** showing.

Finally, comments regarding the voice mode revealed that, due to lack of visual clues, one participant did not know when their partner was ready (Participant 20) and that there was no major difference between the video and voice modes (Participant 12).

Discussion

The aim of the present study was to assess the effect of three SCMC domains on learner oral output as they completed a dialogic task in each environment. Statistical tests revealed that there was a statistically significant difference in participants' output complexity along both measured dimensions. First, regarding the structural complexity measure (i.e., number of words per utterance participants produced between the voice and VR modes), participants produced more words per utterance when completing the voice task. As seen in Yanguas (2010), the lack of visual clues (gestures and other paralinguistic clues) when completing the task in the voice mode could have forced students to rely on their spoken language. Similarly, as seen in York (2019), the cognitive demands of the VR mode could have hindered complex output. Referring to comments, cognitive demands could have been increased due to a lack of familiarity with completing tasks in VR environments, and the use of gestures could have lessened the need for speaking compared to other modes.

The lexical complexity measure (i.e., number of different words spoken) also revealed a statistically significant difference between the video and VR modes. Participants produced a greater volume of different words when completing the VR task in comparison to the video mode. Reasons for this were not immediately evident in participant comments, however, as seen in Xie et al. (2019), the additional stimuli presented by the VR environment could have allowed for an increased awareness of and access to vocabulary.

As for accuracy and fluency measures, no statistically significant differences were found in the data. This echoes York (2019), who also found that task type was more influential than modality on learners' output accuracy and fluency.

Regarding the sub-question "What are the affordances of each domain for language learning," a simple summary follows. Results suggest that the voice mode may be a suitable domain for promoting student fluency, and the VR domain for focusing on vocabulary learning.

Conclusion

This study aimed to uncover the cognitive affordances of three unique SCMC environments. The specific research question was, "Is there a difference in learner oral proficiency when completing interactive tasks within voice, video, and VR-based SCMC modes?" Thirty students (15 dyads) completed a spot-the-difference task in all three modes where their oral proficiency during task performance was measured along complexity, accuracy, and fluency dimensions. Results of this study suggest that the VR mode pushed learners to produce the highest lexical complexity, however, for other measures, no statistically significant differences were found. This suggests that all three modes may be effective in promoting oral proficiency. A negative interpretation of this would be that if the same results can be achieved with simple voice-SCMC, that is, without spending excessive money on VR environments, then teachers have no reason to worry themselves with keeping up with EdTech trends and demand VR systems for their students. However, with the increasing complexity and possibilities that VR environments provide, there is a great potential of this modality to support language learning in future classrooms. The current study utilized very

little of VR's affordances (interactivity, feedback, access to native speakers, etc.) and so our own future studies will focus on incorporating more features of VR and assessing how such features affect language learning from both cognitive and affective perspectives.

The implications of the current study for teachers is that the embodied experience of completing tasks in VR may be both a source of increased engagement, but due to reduced FLA, may also be beneficial in improving learners' oral proficiency. Of course, we are aware that the head-mounted display supported VR environment used in this project is far from being practically implementable in language classrooms, but similar VR systems are being developed and rapidly becoming more portable (see the Oculus Quest for example of a standalone system). Future research should increase the number of task types that participants complete to explore how the VR modality affects performance along the task complexity dimension (see York, 2019). The effect of different VR systems on learner output should also be explored to understand if graphical fidelity and interactivity is an influential component of VR (e.g., smartphone VR, standalone VR sets, desktop PC-enabled VR, etc.).

References

Bygate, M. (2001). Effects of task repetition on the structure and control of oral language. In M. Bygate, P. Skehan, & M. Swain (Eds.), *Researching pedagogic tasks: Second language learning, teaching and testing* (pp. 23–48). Longman.

Bonner, E., & Reinders, H. (2018). Augmented and virtual reality in the language classroom: Practical Ideas. *Teaching English with Technology*, *18*(3), 33–53.

Hawkinson, E., Mehran, P., & Alizadeh, M. (2017). Using MAVR to bring new dimensions to the classroom. *The Language Teacher*, *41*(3), 30–32.

Housen, A., & Kuiken, F. (2009). Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics, 30*(4), 461–473. **https://doi.org/10.1093/applin/amp048**

Lin, H. (2014). Establishing an empirical link between computer-mediated communication (CMC) and SLA: A meta-analysis of the research. *Language Learning & Technology, 18*, 120–147.

Lotherington, H., & Jenson, J. (2011). Teaching multimodal and digital literacy in L2 settings: New literacies, new basics, new pedagogies. *Annual Review of Applied Linguistics*, *31*, 226–246.

Melchor-Couto, S. (2017) Foreign language anxiety levels in Second Life oral interaction. *ReCALL, 29*(1): 99–119. **https://doi.org/10.1017/S0958344016000185**

Melchor-Couto, S. (2018) Virtual world anonymity and foreign language oral interaction. *ReCALL, 30*(2): 232–249. **https://doi.org/10.1017/s0958344017000398**

Ministry of Education, Culture, Sports, Science, and Technology. (2015). On integrated reforms in high school and university education and university entrance examination aimed at realizing a high school and university articulation system appropriate for a new era. **https://www.mext.go.jp/component/english/__icsFiles/afieldfile/2015/03/31/1353908_1.pdf**

Ortega, L. (1999) Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, *21,* 109–148.
Pennycook, A. (2000). English, politics, ideology: From colonial celebration to post-colonial performity. In T. Ricento (Ed.), *Ideology, politics, and language politics: Focus on English* (pp. 107–120). John Benjamins.
Satar, H.M. & Özdener, N. (2008). The effects of synchronous CMC on speaking proficiency and anxiety: Text versus voice chat. *The Modern Language Journal*, *92*(4), 595–613.
Skehan, P. (1991). Individual differences in second language learning. *Studies in Second Language Acquisition*, *13*(2), pp. 275–298.
Skehan, P. (1998). *A cognitive approach to language learning.* Oxford University Press.
Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510–532. **https://doi.org/10.1093/applin/amp047**
Skehan, P., & Foster, P. (1997). The influence of planning and post-task activities on accuracy and complexity in task based learning. *Language Teaching Research, 1*(3), 185–211.
Su, Y.C. (2016). The international status of English for intercultural understanding in Taiwan's high school EFL textbooks. *Asia Pacific Journal of Education, 36*(3), 390–408. **https://doi.org/10.1080/02188791.2014.959469**
Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. *CALICO Journal, 13*(2), 7–26.
Wigham, C.R., & Chanier, T. (2015) Interactions between text chat and audio modalities for L2 communication and feedback in the synthetic world Second Life. *Computer Assisted Language Learning*, *28*(3): 260–283. **https://doi.org/10.1080/09588221.2013.851702**
Xie, Y., Chen, Y., & Ryder, L.H. (2019). Effects of using mobile-based virtual reality on Chinese L2 students' oral proficiency. *Computer Assisted Language Learning*, 1–21. **https://doi.org/10.1080/09588221.2019.1604551**
Yanguas, I. (2010). Oral computer-mediated interaction between L2 learners: It's about time! *Language Learning & Technology, 14*(3), 72–93.
Yanguas, I., & Bergin, T. (2018). Focus on form in task-based L2 oral computer-mediated communication. *Language Learning & Technology, 22*(3), 65–81. **https://doi.org/10125/44657/**
York, J. (2019) Language learning in complex virtual worlds: Effects of modality and task complexity on oral performance between virtual world and face-to-face tasks. [Doctoral dissertation, University of Leicester]. **https://leicester.figshare.com/account/articles/10242452**
York, J., Shibata, K., Tokutake, H., & Nakayama, H. (2021). Effect of SCMC on foreign language anxiety and learning experience: A comparison of voice, video, and VR-based oral interaction. *ReCALL*, *33*(1), 49–70. **https://doi.org/10.1017/S0958344020000154**

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*(1), 1–27. **https://doi.org/10.1093/applin/24.1.1**

Ziegler, N. (2016). Synchronous computer-mediated communication and interaction. *Studies in Second Language Acquisition, 38*(3), pp. 553–586. **http://doi.org/10.1017/S027226311500025X**

Appendix A

Pre-Task worksheet

Spot the difference task: Warm up

In this lesson, you will work with a partner. You will compare two pictures of the same room. There are six items placed in the room, but some of them are in a different place to your partner. This is a spot-the-difference activity. You will have to decide which of the items are the same and which are different.

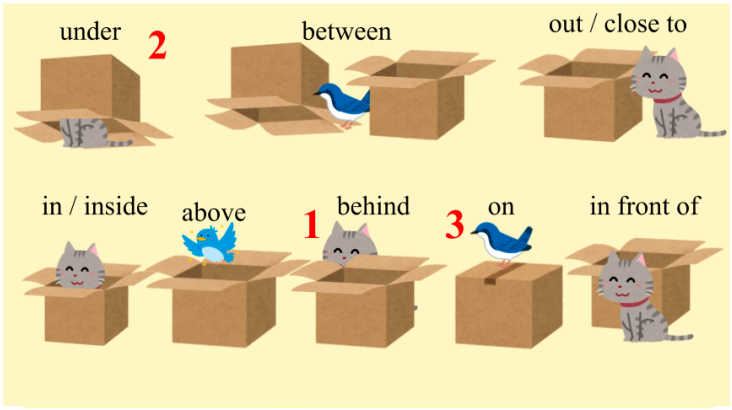
Activity 1: What is in your living room?

Think of the objects that you find in a living room and make a list of 10 things here:

_____	_____
_____	_____
_____	_____
_____	_____
_____	_____

Activity 2: Positions of place

Look at the following picture. The cat 🐱 and bird 🐦 are in different positions in relation to the box.

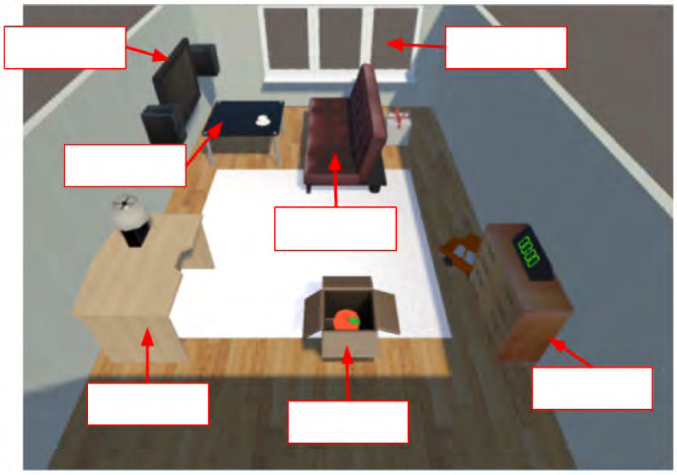


Please write sentences for the three pictures labelled **1**, **2** and **3**. For example: *The cat is in the box.*

1. _____
2. _____
3. _____

Activity 3: Label the items.

Please label the items marked with a textbox:



Activity 4: Where are the objects?

In this room, there are the following objects:

 Apple	 Lamp	 Clock
 Car	 Present	 Cup

Please mark the following sentences correct or incorrect.
Please use a ✓ for correct sentences.
Please use a ✗ for incorrect sentences.
For example:

The apple is on the box. ____X_____

1. The clock is on the chest of drawers. _____
2. The car is in front of the chest of drawers. _____
3. The present is on the sofa. _____
4. The lamp is on the desk. _____
5. The cup is in front of the tv. _____

Authors' Bios

Hayato Tokutake is a Master's student studying information sciences at Tokyo Denki University. He conducts research on the application of virtual reality in language education.

James York is a lecturer at Tokyo Denki University where he researches the application of games and play for teaching languages. He also co-edits *Ludic Language Pedagogy,* an open-access journal.

Hiroshi Nakayama is a professor of instructional technology at Tokyo Denki University, conducting research on the integration and application of technology for educational purposes.