Title	Improving tandem language learning using immersive 360° video
Sub Title	
Author	Kitsing, Inga Kunze, Kai
Publisher	慶應義塾大学大学院メディアデザイン研究科
Publication year	2022
Jtitle	
JaLC DOI	
Abstract	
Notes	修士学位論文. 2022年度メディアデザイン学 第946号
Genre	Thesis or Dissertation
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO40001001-00002022-0946

Master's Thesis
Academic Year 2022

Improving Tandem Language Learning Using Immersive 360° Video

Keio University
Graduate School of Media Design

Inga Kitsing

A Master's Thesis
submitted to Keio University Graduate School of Media Design
in partial fulfillment of the requirements for the degree of
Master of Media Design

Inga Kitsing

Master's Thesis Advisory Committee:

Professor Kai Kunze	(Main Research Supervisor)
Project Senior Assistant Professor Yun Suen Pai	(Sub Research Supervisor)

Master's Thesis Review Committee:

Professor Kai Kunze	(Chair)
Project Senior Assistant Professor Yun Suen Pai	(Co-Reviewer)
Professor Nanako Ishido	(Co-Reviewer)

Abstract of Master's Thesis of Academic Year 2022

Improving Tandem Language Learning Using Immersive 360 ° Video

Category: Design

Summary

Virtual Reality (VR) has been actively alternating learning opportunities lately and yet has the potential to further expand in the upcoming years. 360-degree cameras combined with VR techniques towards more immersive environment has little studies, especially for language learning domain. This paper presents the concept of an immersive experience that integrates Virtual Reality (VR), sound for visualizing, sharing and engaging with a tandem partner's environment through VR headset connected to a 360-degree camera for in situ language learning. It also attempts to understand its effectiveness in contrast with traditional learning.

Keywords:

VR, first-person view, language learning, 360 camera, contextual learning

Keio University Graduate School of Media Design

Inga Kitsing

Contents

List of Figures

Acknowledgements

I am indebt to Professor Kai Kunze, Professor Yun Suen Pai for continuous support and the whole 2020 batch for helping me out with the user study. Your valuable feedback is what made this prototype develop and become more engaging day by day. I wish to show my appreciation to George, Annika, Cady, Sakurai, Kinga, Chen, Cendika, Pan, Flanny for assisting my user study at every step of the way.

I am absolutely grateful for this 2 years experience at KMD, exploring the field of Interactive Media Technology and HCI, which would not have been possible without the support of MEXT grant. I wish to show my appreciation to Elvira's help, without who finalizing my data and reorganising the logical structure would have been impossible. I would also like to thank Keio-sensei, Kouta Minamizawa and Sohei Wakisaka for the final remarks of my write up. Finally, a special shoutout to Geist lab, Project Room SA team, SDM cozy office rooms and Lawson that would not let us starve late nights at Kyoseikan.

Chapter 1
Introduction

1.1. Background

Many people are interested in learning a new language, but are discouraged by how time consuming it is [8]. Despite the opportunities a foreign language offers, it is especially challenging to keep going in a persistent manner after reaching a certain level of fluency since learners do not see their progress a rapidly as it was at an elementary level. L2 (second language) learners often times find it to find motivation for studying and tedious to regularly sit down and engage with educational material.

According to the The Foreign Service Institute (FSI) of the United States Department of State projects how long it usually takes to master a foreign language based on 70 years of continuous teaching [9]. Because this research is focusing on Japanese language acquisition, let us take a look at the relevant data regarding this Asian language. As stated in FSI's report, Japanese is characterized as a category IV language ("super hard", as they are exceptionally difficult for native English speakers) and will take up to 88 weeks or 2200 class hours to reach a good level of proficiency. Many Japanese learners just lack such time commitment or cannot maintain their motivation levels to devote a significant amount of time and resources to intensive education.

The majority of language learning classrooms prioritize communication proficiency when teaching. Four fundamental language skills—speaking, listening, reading, and writing — are traditionally included in linguistic learning objectives for language learning intended to better divide progression and various levels according to language proficiency. The linked skills of vocabulary and grammar, which include spelling, syntax, and pronunciation, are used by language learners as instruments to build sentences messages. Grammar and vocabulary knowledge

are important for a student to efficiently construct and deliver their thoughts.

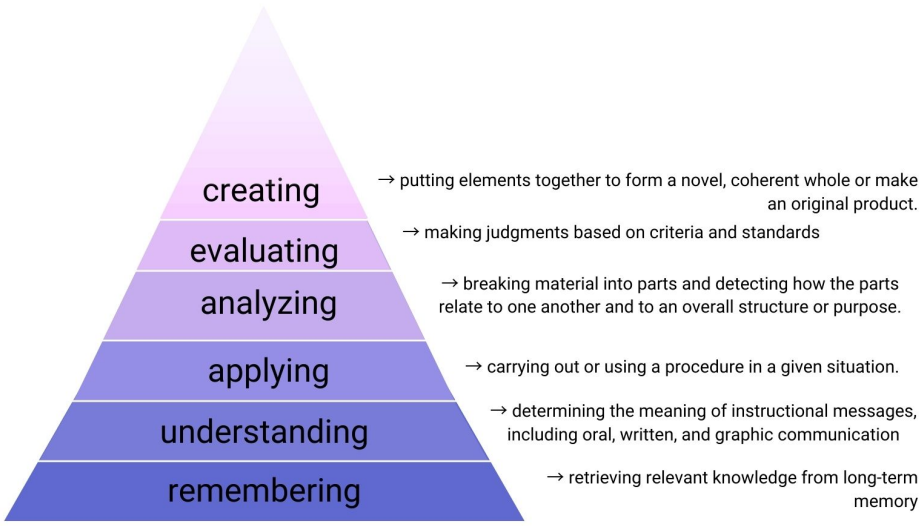

Figure 1.1 Bloom's taxonomy in accordance with corresponding linguistic abilities. Based on Bloom's taxonomy in 1956 [1].

Bloom's Taxonomy (see figure 1.1), is a hierarchical ranking of learners' cognitive abilities language most educators use in order to generate learning goals for lesson plans. Seven of the taxonomy's categories such as remembering, understanding, applying, analyzing, evaluating and creating can be applied accordingly to improve a language acquisition method. Bloom's Taxonomy is less of a hierarchy when it comes to language learning and more of an interdependent set of skills that are constantly reinforced through language practice. Understanding how VR technologies can support learners in developing the four fundamental language skills (speaking, listening, reading, and writing) as well as related abilities (grammar and vocabulary) can build up complete functional learning curricula [3].

Some researchers claim that reading texts when learning a second language is

expected to provide essential content that is vital for mastering all aspects and therefore reading is considered to be the main vocabulary acquisition way in L2 development [10].

1.2. Can We Learn Faster?

Learning languages by getting closer its native speakers and immersing in their environments is assumed to be an re-enforced way to add context and speed up the language learning. However, it's obvious that adults just can't learn "implicitly" the way small kids do, by being surrounded by native speakers all day. Unfortunately grown ups rarely can afford the luxury to go to another country financially or limited by the time. Some of the widely known ways to learn a language while achieving the progress quickly is, for instance, studying abroad, using apps like Duolingo or Rosetta Stone, or even having one-to-one lessons with your foreign partner.There are quiet a few attempts to modernize the classical learning model, especially with the use of technology.

Speaking practice is crucial for the growth of a second language (Gass, 2018), but there are rarely many opportunities for in-person speaking practice in the classroom. Utilizing technology to engage students in online collaboration projects (for example projects that utilize digital technologies to connect students in geographically dispersed classes) or one-on-one discussions with subject matter experts on specialized platforms is one way to address this issue.

Most importantly though, cultural involvement perhaps plays a deciding role for reinforcing cognitive abilities for mastering language proficiency in the context of immersive language learning. It is indeed an uneasy task to learn a new language when being an adult and learn new vocabulary [8].

Conventional methods of learning often involves cramming grammar, translating texts and vocabulary repetition is frequently used to expose learners to words outside of their native language. Usually traditional learning relies on printed materials such as textbooks, dictionaries and learning content from a blackboard while sitting at a desk. Listening skill is taught by prerecorded audio materials that are often coming not from a native speaker. Collectively, this gives the learner that necessary skill for a basic interaction in a foreign language.

This methodology has certainly proved to be effective by centuries and that is why not by chance is called traditional way of learning. However it might block learner's understanding of the native speech, thinking of the words to say or thinking too much that will result in falling behind. Another negative aspect that is very hard to fix at a later stage is acquiring a wrong accent, that might be crucial for some people that work hard to integrate in the foreign environment in order to escape any language barriers.

1.3. Virtual Reality As a Learning Aid Of the 21st Century

Virtual Reality (VR) is a simulated a computer-generated environment with realistic-looking scenes and several levels of immersion (semi-immersed, non-immersed and fully-immersed). It is a kind of technology, that is capable of immersing the user in a different surroundings from the ones they are at the moment. This environment is perceived by using a Virtual Reality headset or helmet.

Figure 1.2 Virtual Reality headsets: Oculus Quest (black) and Oculus Quest 2 (white, new generation) by Meta.

The creators and promoters of VR headsets assert that they could give users

the opportunity to experience real-world instruction without having to leave the classroom.

Recently there has been seen an emerging interest in applying 360 videos combined with virtual reality techniques towards more immersive environments, so that a user can directly experience and interact with its content. For this project, immersive content is shared between the two users and applied towards enhancing foreign language learning. In this short paper we focus on the use of immersive VR in the context of foreign language learning and answer whether the immersive context (video/images) fitting to the discussion topics can enhance the learning compared to a VR environment unrelated to the conversation content.

Among the various types of user interactions designs with the VR hardware, the interest of adding a 360 camera in immersive VR technology for education seems to be quite high, however the review results show that mostly these technologies remain in an experimental state and have not been implemented yet, mostly due to its poor performance [11].

1.4. Challenges

Being called a “learning aid of the 21st century” (Rogers, 2019), VR technology offers an extra element to the traditional learning and teaching. Radianti [11] reviewed 59 peer-reviewed articles published between 2009 and 2018 that included virtual reality in the context of learning and found that VR applications for education were most frequently used to teach practical knowledge (33%), declarative learning (25%) and problem-solving skills (12%) the rest was communication and soft skills (10%), of which only 2% were specifically for learning a language. That is surprising, because virtual worlds have great educational advantages and potential impacts on language learning, such as providing location-specific and socially interactive learning, including but not limited to field trips, city tours, role playing.

There have been certain attempts to apply virtual reality in foreign language learning [3] [5] [12] [6] and other researchers. However, those studies employed VR environments in a pre-recorded manner 1.4 1.3. Those applications normally use computer-generated virtual space that was developed to be as close to realistic

Figure 1.3 Related works: A screenshot from a 3D virtual learning platform called Mondly [2].

Figure 1.4 Related works: A screenshot from a 3D virtual learning platform called Mondly [2].

as possible, because of how immersive that experience might be [6] [11]. These environments may give the learner a feeling of presence and thus this environment might in turn positively affect improving communication skills [6]. The degree of immersion is seen to be crucial to distinguish in order to be able to consider the learner's perspective and whether they prefer high immersion or lower immersion levels. A key element that enhances the authenticity of the learning experience is in fact using a VR headset compared to consuming educational content on a 2D display. Therefore we would like to underline the importance of distinguishing between several levels of immersion ref when it comes to learning using virtual reality environments. The levels of immersion provided by various immersive media implementations affect not only the user experience of learners, but also their learning outcomes [3].

Getting this far, it is already obvious that existing research and applications as service that exist in the market are rather in the beginning of the immersion progression scale, because it (1) has a strict scenario to follow that was pre-generated by the developer therefore the level of interaction is limited to a in-program character that has a limited number of pre-coded answers or interaction scenarios, (2) it is 3d animated environment, that, of course is immersive but it is far from being realistic looking unless it is high quality 3d animation, that requires a lot of resources, whereas learning is rather a creative process and teachers are often encouraged to be more flexible in their interactions with their students. While we can train an AI to develop numerous scenarios and ways to rephrase and add fluency in speech, she will still not be a human. Only humans have this pedagogically sound language learning approach. On top of that, while online studying has been researched as one-on-one foreign language interaction and fewer studies has been done in regards with virtual educational prerecorded experiences, real life footage being filmed on 360 degree camera while broadcasted life onto a Virtual Reality headset for in the frames of language instruction is lacking. Therefore, we believe that VR x 360 degree camera can potentially modernize the dynamics of tandem language learning experiences.

1.5. Proposal

As a part of interactive 360 ° experience, the panoramas, videos, text and narration are providing a feeling of being in a real location for the user. In regards with learning, users are able to explore a location without physically going there. Users can look in any direction and navigate around while interacting with the images, videos, and information embedded in the simulation thanks to the 360 ° footage. The benefit of 360 ° experience is that it provides users with a safe, virtual experience at a relatively low cost.

The purpose of the study is to explore whether immersive content in the real world (shot on 360 degree camera) broadcasted virtually using a VR-headset can enhance foreign language learning compared to conventional learning methods, where environments usually are not related to those of the topic of conversation.

Figure 1.5 Originally proposed concept of the system that incorporates both 360 degree footage and a VR headset demonstrating the variety of 360 ° cameras.

Expected outcomes include expect that remote immersive 360-degree tandems can positively affect foreign language learning by expanding the vocabulary, life

interaction and physical movement.

Therefore the issue with existing learning language tools is that the immersion level might be good, but the user is not interacting enough while being "presently there". To make the learning experience more authentic, VR technology has demonstrated to have that notion of learning knowledge and speaking skills in context as they are not only mirroring the real life setting, but literally immerse the user there. Thus, this research looks into conversation practice in VR.

1.5.1 Work Flow

First, desk research was carried out on what to focus on while testing the system in order to further state the significance of contextual learning and context being more important than the actual content. Second, the actual VR + 360 camera system is designed in accordance with the idea of tandem partner environment exchange. After that the implementation stage is to actually carry out the pilot experiment (i.e., museum, field trip, restaurant) and finally evaluating the effectiveness of student learning.

1.6. Thesis Structure

This thesis consists of 5 chapters. Chapter 1 presents the background of the problem and underlines the need of further in-depth research in the current field. Following chapter 2 that investigates relevant research in the field of VR language learning and whether there were attempts to investigate how effective that was. Chapter 3 outlines the concept that was proposed for the system and in turn its implementation is further looked into in Chapter 4. Finally, Chapter 5 provides the summary of research and discusses some possible implications and of this research.

Chapter 2
Related Works

2.1. Virtual Reality in Education

Studies that look into various approaches for enhancing language learning have dramatically increased over the past decade. Suggested innovative techniques claim to provide more engaging learning experience to learners and, in turn, help them to store and retrieve vocabulary in a more efficient way. Currently, only a small percentage of studies (27 out of 167) focus on language learning, and sample sizes are typically small [13].

2.2. Levels Of Immersion

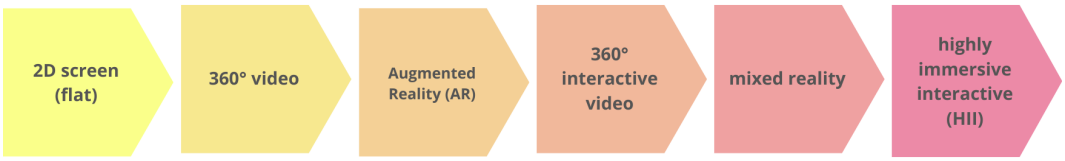

Figure 2.1 Levels of immersion according to Hayes [3].

As seen on the scale above 2.1, Aleshia Taylor Hayes provides an overview of the levels of immersion that immersive media tools offers from immersion perspective, progressing from 2D flat screen (1 - the least immersive approach), 360 degree video, Augmented Reality (AR), 360 Interactive Video, Mixed reality and, finally, highly immersive interactive (HII) (6 - that is considered to be the most immersive nowadays). Those levels of immersion, in fact, influence not only learners' user experience, but also their outcomes after being taught [3].

2.2.1 Immersion Off a Flat Screen

When it comes to virtual experiences on a flat screen (also known as a 2.5D immersive environment), users gain a sense of immersion through control and, in some cases, the creation of their own personal avatars. While they can be represented as 360-degree videos, they actually provide a panoramic view with a mouse, rather than being fully immersed [14].

2.2.2 Augmented Reality (AR)

In case of Augmented Reality (AR), it establishes digital information on top of physical objects or environments. According to some researchers [15], three major characteristics of AR can be identified: (1) combines the real and the virtual, (2) is real-time interactive, and (3) is registered in three dimensions. This technology is most commonly found in mobile device applications that use the smartphone's camera to scan the environment and generate augmented elements in it, making it extremely accessible and/or affordable in an educational setting.

2.2.3 360° Video

In terms of 360 video, these cameras can be viewed not only on a flat monitor, but also through a VR headset as immersive video, allowing users to turn their heads around and enjoy the 360 degree experience. When users wear the goggles, the surrounding reality is replaced by visual representation, creating a stronger sense of immersion. Since these users are isolated from the physical world, and the images can be displayed in omni directional way, viewers see this environment as the fully immersive one, thus enhancing the overall experience and making it more enjoyable.

2.2.4 Mixed Reality (MR)

Following, Mixed reality is seen as the next wave in computing. It liberates us from screen constraints, allowing us to use instinctive interactions with data at home and with our friends. This type of reality basically generates all sorts of

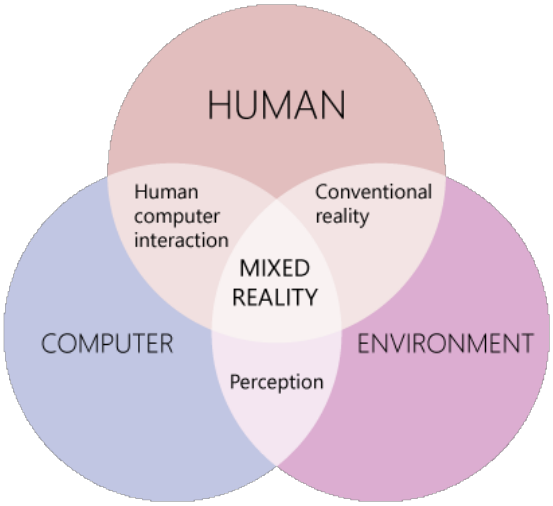

Figure 2.2 Interaction among computers, humans, and environments in the context of Mixed Reality (MR) [4].

interactions by providing truly stunning holographic representations of people as well as highly accurate holographic 3D models [4] – see figure 2.2.

2.3. Sociolocultural theory

The theory proposed by Vygotsky contradicts the theory of the French psychologist Ribot, who argued that attention is associated with emotions and is caused only by them. He confidently emphasized the relationship between voluntary attention and emotions, suggesting that the activity and duration of attention directly depend on the intensity and duration of emotional states. Ribot substantiated the driving theory of attention, which states that the main role in the processes of attention is given to movement, and not speech, as Vygotsky notes.

Vygotsky's sociocultural theory [16] is where contextualized learning originates. The interaction of interpersonal, cultural-historical, and individual factors on learning and development is emphasized in this theory. Furthermore, the cultural factor demonstrates that learning and development cannot be separated from its

contexts. A context should include the environment, objects, and people, as well as the interaction between them.

2.4. Telexistence and Telepresence

Telepresence and telexistence research focuses broadcasting a person's wholly existence or presence to distantly connected sites. The challenge that researchers are focusing on is usually related to creating a deeper sense of presence for distantly connected users beyond the traditional video communication platforms. Several researchers have contributed to developing unique mobile robotic systems for transmitting gestures [17], facial expressions and physical motions [18] and [19] [20].

In particular, some of them are concentrating on developing wearable devices such as T-leap - a one to many telepresence system that incorporates 360 degree camera and casts it to remote locations [21], "Polly" [22], which is a shoulder mounted telepresence device that alters the direction of the view to explore the surroundings. In addition, certain projects are implementing a remote avatar that may be manipulated remotely by the ghost user, such as *Teroos* [23] or *Gusty Avatar* [23], *MH-2* that by indicating direction and timing realizes telepresence of several distantly connected people [24].

2.4.1 Improving Skills in a Remote Environment

Telepresence provides the user the sense of being present near the faraway partner, and co-presence is the concept in which the other person is perceived as being there locally with them. w the people can share and improve human skills by sharing immersive experience with real world context [25].

Previous research in remote collaboration has investigated several methods for sharing visual space via video streaming using cameras. Wearable cameras are one of the key techniques for sharing first-person visual information since they allow us to capture and transmit a first-person view. Precisely, Jack in Head jack in head [26], that allows remote users to explore the scene independently of the local user. Interestingly, this system was designed to follow the direction of the

head of the active user, and before being broadcasted to the local user the system performs motion stabilization (for omnidirectional video footage).

Furthermore, there has been works on sharing body functions - Nin Nin robot, for instance offers the concept of "lending" a part of the body, would it be an ear, an eye or a leg throughout society [27]. NIN NIN allows to remotely share physical functions such as sight and hearing by connecting to via the Internet from a smartphone or tablet. For example, to share your eyesight with a visually impaired person, or a speaker to share your lips with someone who does not speak English in order to collaborate.

In addition, Fusion by Embodied Media laboratory at Keio University in the frames of Fusion project have implemented is a telecollaboration technology that allows two people to operate and collaborate remotely by sharing the same point of view and physical area [28]. It provides three levels of communication: direct movements utilizing humanoid hands, enforced postures by holding and moving surrogate hands, and induced motions by moving surrogate hands beyond physical reach enforcing real body communication.

2.5. Existing Research

Finding opportunities to practice a language in real-world circumstances is one of the most difficult aspects of language learning.

2.5.1 The Effects of Virtual Reality Learning Environment on Student Cognitive and Linguistic Development

Chen [29] in their study evaluated the VR language learning effectiveness based on Bloom ' s cognitive complexity level based on virtual worlds, cognitive complexity and contextuality. Particular features such as immersion, ease of use and help seeking are proved to positively affect perceived cognition. Ease of use is essentially based on clarity: easy operation, clear instruction.

Virtual reality technologies are already enhancing the realism and smoothness, and Chen ' s experiment (2016) participants have reported to " learn better when they have the ability to change the view position ". Therefore, designing a similar

system equipped with an actual 360 ° camera is reflected to significantly boost students ' concentration and motivation level. Finally, help seeking was identified as the most crucial part in the whole test: having a teaching assistant ready to provide appropriate advice in both technical issues and English learning exceptionally helped the learning [30].

Context is particularly interesting, because the role of context in learning is much more important than the content itself. It is reflected to guide the learners to develop insights through balanced, organic and successful environments. The meaning of a word does not lie in itself, but in its use. Among the measured types (phonological, morphological, vocabulary, grammar and syntax), the most significant difference was found in the scores for identifying and matching vocabulary to sentences correctly, then the grammar and syntax as the participants performed very well on completing coherent sentences with given phrases.

2.5.2 Implementing VR-Enhanced Tasks in Chinese Language Learning

Tseng in their study investigated to what extent immersive experiences created by VR technology enhance language learning and if so, to what extent. The study was conducting among 16 students of Southeastern university taking Elementary Chinese course, who were mostly from 18 to 22 years old and a mean of 19.7 years old. Their Chinese language proficiency ranged from elementary to pre-intermediate level and the tasks they had to do within the experiment was developed according to the main course curriculum.

Researchers chose Unity, Google Play, and the matching 3D Google Cardboard tools in relation to VR tools after experimenting with more than ten VR tools and analyzing their pedagogical applications. As for learning content, it was mostly filmed on cellphones and professional 360 degree camera. Then the users were asked to undergo several close to real life scenarios where they could use their Chinese skills. Overall, this study revealed students' perceptions and attitudes toward four communicative tasks made possible by two types of interactive VR tools and apps, Unity and Google Cardboard.

Tseng and Geng discovered that when they used Unity to incorporate tasks in

the media lab, only one pair of students performed the task at a time, with the instructor and their classmates observing on the spot, giving the instructor more opportunities to fully support each acting pair and attend to their here and now needs. The current study also included a very limited set of linguistic structures and functions, and expanding beyond this set is an additional area for future research.

However, this proved to be a disadvantage of the Google Cardboard device, which they used in a regular classroom. As a result, not every student received full support and attention from the teacher. However, both systems have demonstrated the ability to empower and boost learners' confidence in speaking.

2.5.3 VR for Teaching ESP Vocabulary

A more accurate study had been conducted in regards with English for Specific Purpose (ESP), that not only confirmed the recent growing interest in wearable devices globally, but also examined whether virtual reality headsets help to retain vocabulary related to their field or not [5]. The experiment lasted for about 6 weeks and participants were interacting with virtual objects through watching 360 degree videos related to the focus vocabulary. They learnt such words as depression, addiction, suicide, etc (50 vocabulary items in total) after applying the VR headsets. Research designers used the pre-recorded videos for the experiment related to the topic of the class. The procedure did not include any prior teaching, and remarkably the participants had a range of different scores for English as a second language exam, meaning the students had various level of English.

In general, the study revealed a (P 0.001) rejection of the null hypothesis and a mean difference between pre and posttests. As a result, the mean of the post intervention test is statistically significantly higher than the mean of the pre intervention test – see figure 2.3.

2.5.4 One-On-One Foreign Language Speaking Practice in VR

Another study looked into acquisition of German as a second language and conducted a number tests with one person, not a typical one test vs several users.

		Mean	N	Std. Deviation	Std. Error Mean	C.V	*P Value*
Pair 1	Pretest	32.000	20	8.0328	1.7962	25.1%	-
	Posttest	39.800	20	6.5261	1.4593	16.4%	0.001

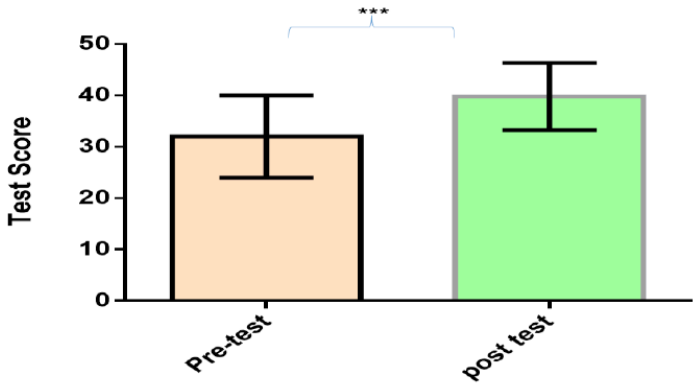

Figure 2.3 Related works: Madini's result for testing Saudi female students' English for specific purposes vocabulary retention demonstrating significant improvements at the post-test results [5].

It investigated the dynamics of five one-on-one virtual reality tutoring sessions. A CEFR B1 (intermediate) level of German student with the help of the tutor practised their foreign language speaking skills within the computer generated VR system. They based their research among others on perception and learning preference of the student. The adult learner of German language was also classified as a willing to communicate outside the classroom student, that also added up to the success of the current experiment [6].

Significantly, researchers used the VR application that was pre modelled and pre programmed VR animations using vTime online platform. The authenticity, scaffolding of the outside environment, all contributed to proving the hypothesis that VR settings can spark students' creativity. On top of that, after analyzing both the tutors and the students bullet journals, it was found out that it is indeed an immersive way to study languages in a format of tandem learning despite the fact it is computer generated and rather low quality images.

This study found that social VR applications can provide pedagogically beneficial environments for the development of foreign language speaking skills.

2.6. Summary

Currently, research on VR and immersive language learning has been in its embryonic stage in terms of scope, focus, and especially the variety of languages investigated. However, the majority of studies on VR and language education found that VR tools had a positive impact on learners' motivation. In addition, applying virtual environments as a means of improving learning process has shown significant advantage [12].

Getting this far, it is already obvious that existing research and applications as service that exist in the market are rather in the beginning of the immersion progression scale, because it (1) has a strict scenario to follow that was pre-generated by the developer therefore the level of interaction is limited to a in-program character that has a limited number of pre-coded answers or interaction scenarios, (2) it is 3d animated environment, that, of course is immersive but it is far from being realistic looking unless it is high quality 3d animation, that requires a lot of resources, whereas learning is rather a creative process and teachers are

Figure 2.4 Related works: Conversation setting using vTime online pre generated VR environment for tandem conversation purposes by Regina Kaplan-Rakowski and Alice Gruber [6]

often encouraged to be more flexible in their interactions with their students. While we can train an AI to develop numerous scenarios and ways to rephrase and add fluency in speech, she will still not be a human. Only humans have this pedagogically sound language learning approach.

On top of that, while online studying has been researched as one-on-one foreign language interaction and fewer studies has been done in regards with virtual educational prerecorded experiences, real life footage being filmed on 360 degree camera while broadcasted life onto a Virtual Reality headset for in the frames of language instruction is lacking. Therefore, we believe that VR x 360 degree camera can potentially modernize the dynamics of tandem language learning experiences.

Several critical points that were revealed during the literature review were:

1. Choice of technology implemented for the study can be crucial and can jeopardize the study flow.

2. VR tools are ubiquitous nowadays in the scope of applying it in field training, however incorporating Virtual Reality for language learning is still in embryonic stage.

3. Although extensive research has been conducted on the impact of videos on English as a second language, much less is known about the impact of new immersive videos or 360 degree videos on vocabulary retention, not to mention less common languages for learning.

4. Adequate statistical information is crucial in order to see the effect size of the new VR tool [31].

5. Maturation, test effects, statistical regression to the mean (RTM), and contemporaneous effects of innovations in practice are some methodological concerns that may affect the results of such design [32].

2.7. 360 degree video

360-degree video is an immersive video format in which the camera captures a full sphere of video (omnidirectional camera or a set of different cameras), allowing the audience to view the footage from every direction: look up, down, and turn

around in a circle, allowing them to feel immersed in the environment shown in the video. 360 ° videos are best viewed with a VR headset, such as an Oculus or a Google Cardboard, which is a low-cost solution that uses your smartphone as a viewer. 360 video appears to be everywhere these days, including advertising and content marketing. Although it is sometimes referred to as virtual reality (VR), 360 video differs in that you cannot reach out and manipulate objects within your field of view.

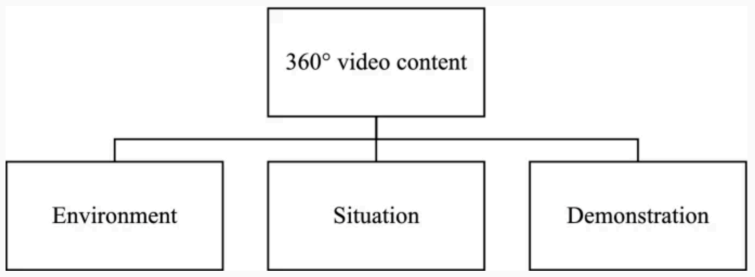

Figure 2.5 Three types of 360 degree videos within the video content.

When the contents of the 360-degree videos used in the dataset were examined, three categories emerged, as shown in Figure 2.5. All of the videos in the dataset can be classified into one of three categories. First, the *environment* videos depict a natural phenomenon like a rainforest or some sort of historical site. Second, the emphasis in *situation* videos is on specific and spontaneous human behavior in the above spoken environment. People interact with their surroundings in a non scripted manner while behaving unpredictable, and this is the focus of the video. Third, *demonstration* videos. These videos, in contrast to the previous two categories, are predetermined as they are designed to carry out a specific operation or action.

Interestingly, as shown in figure 2.5 and figure , 50 percent of the videos in the dataset are *situation* videos. Furthermore, some videos focus on human behavior in specific contexts, such as [33], who show human interactions in a local neighborhood, and those that show construction site scenarios [7]. Finally, in this category, there are two studies that used videos of human actions in a specific situation. A certain researcher had used a video of a mountain climber in Nepal [34], while another demonstrated driving behavior in a dangerous traffic

situation [35] [36].

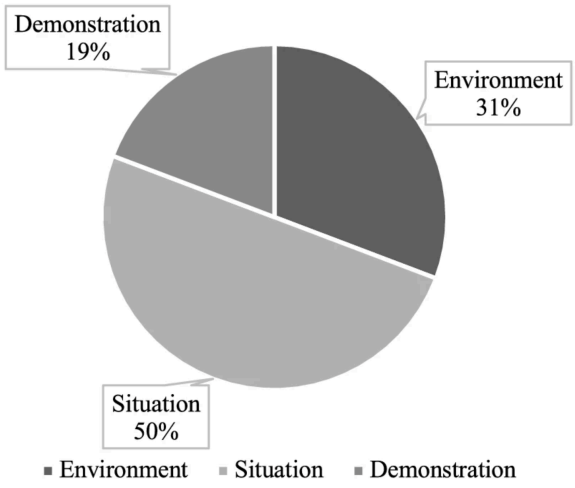

Figure 2.6 Three types of 360 degree videos within the video content by category retrieved from Evans [7]

Chapter 3
Concept Design

3.1. Concept

3.1.1 An Immersive Real-Time Language Learning System

Currently the amount of various sorts of language learning helping applications is ubiquitous. Flashcards, online lessons, learning through subtitles and watching news in a foreign language, specialized language schools in the country where locals use that language as their native one. However, all of them commonly lack one important component – that is context. Until now, there are no well-developed and most importantly proven guidelines or principles for incorporating VR tools and apps into the language classroom in the literature on second language acquisition.

The current state in the context of Virtual Reality is exceedingly focusing on contextualized learning. These include active learner participation, authentic contexts, as well as meaningful and social interaction [37]. Virtual Reality technology seems to be a perfect medium for combining them together, however creating an authentic learning experience while keeping it interactive, meaningful and engaging at once is an uneasy task.

As a researcher, I would like to propose a system that would improve existing systems for learning languages while providing a high quality teaching ensuring the pedagogical standards together with effectiveness of a tandem lesson. The system connects a "ghost user" that is staying at home, perhaps in another country and an "active user" that broadcasts their environment in real time using a 360 degree camera over the Internet. Thus, the best features of VR system together with one-on-one learning are combined together providing the best authentic, immersive learning experience.

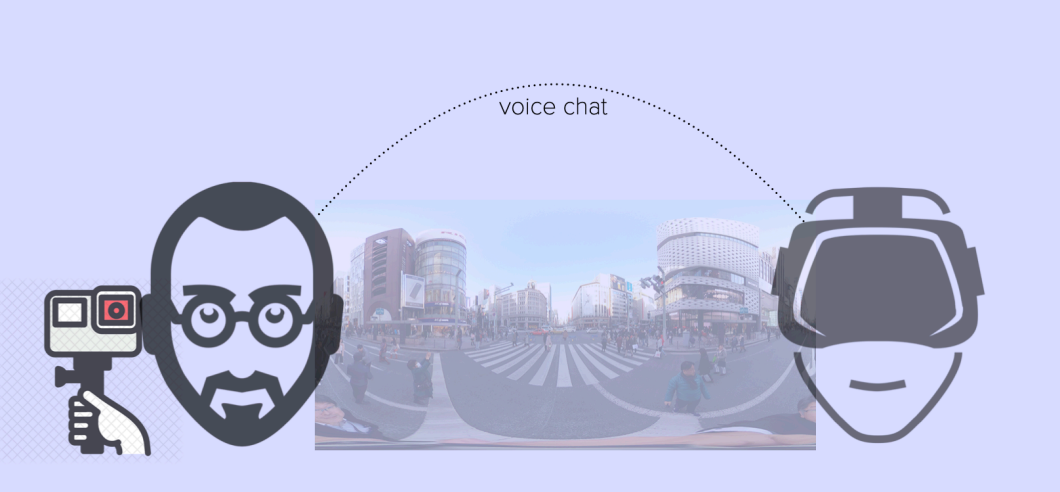

Figure 3.1 Ideation: Concept design prototype for improved tandem language learning experience.

3.2. Research Goals

According to comprehensive desk research carried out earlier and literature review findings regarding the current technologies focusing on providing contextual foreign language learning, it is evident that there is a huge gap in applying immersive technology in the context of language learning.

The goal of this experiment is to find out if using an improved VR set with a 360-degree camera can actually further enhance foreign language learning given its smooth, real immersion for language environment in comparison with VR only carried out tests and possibly active adoption of immersive VR technology combined with a 360 camera.

This thesis therefore aims to solve this problem by introducing a new factor into online language learning, even if it is through VR and that is the real time conversation in 360 degree video by:

- Exploring the design process of an engaging yet meaningful virtual learning system
- Presenting the most pertinent findings from the learners' experiences about

the authenticity and effectiveness of Japanese language learning settings in situ;

- Discussing the implications and limitations of using virtual reality applications for language learning
- Finally, creating a working prototype that takes advantage of the system using a VR headset.

3.3. Research Direction

This study is not purely descriptive but also presents practical application of suggest system design. Applied evaluation includes two focus groups: an experimental group and a control group to compare learning in two different contexts – a conventional learning context without VR technology and a newly created learning context with VR technology. Both groups have the same curricular and pedagogical goals. The results of this comparison will be used to improve experimental design and deepen our understanding of the roles that VR technology can play in various dynamics and settings. We are hoping that this system would perform as a powerful learning tool for any foreign language, although in this study we focus more on Japanese language acquisition as a foreign language.

As described in Chapter 2, researchers have proved that the context is most crucial when it comes to learning. The environment, its objects, and people, as well as interaction among three, must all be included in a context. We are hence hoping to create an engaging narrative for the teaching tandem partner to follow acknowledging the importance of cultural, historical and individual factors.

3.4. Target Users

This study is focusing on individuals that are currently interested in mastering their Japanese skills, especially those that look into ways to improve their conversational abilities. Because our research team is based in Yokohama Japan, we are physically limited to those learners, that are already in Japan. Therefore our

main target users that volunteered to participate in the user study are 21 learners of Japanese coming from different academic backgrounds and those that are currently young working adults or student at Keio University – see figure 3.2.

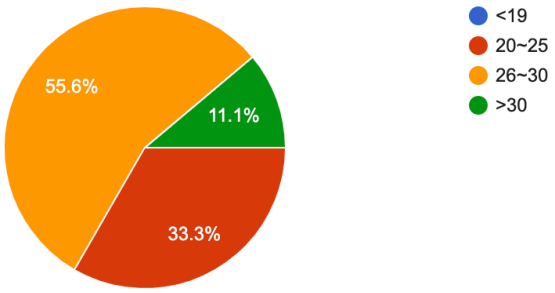

Figure 3.2 Total of 21 participants among 2 user tests and their age distribution. Includes participants from both pilot user group (5 users) and the final user study group (8 users + 8 users in total.

The system essentially is focusing on improving one-on-one conversation practice in the real world circumstances and therefore requires two users. We call them a "ghost user" (primary learner) and an "active user" (in situ language partner that is a native speaker and tutor of Japanese). The active user is located in Japan and broadcasting local environment with the help of 360-degree camera while the ghost user has this footage put forth on their Virtual Reality headset - see figure 3.3 3.4.

We are developing students ' language proficiency in two functions: describing the space and physical settings (1) and communicating about the physical movements of a person and/or moving objects in the frames of a particular topic (scenario) (2).

Essentially, it is a two-way interaction with a Japanese friend (experiment assistant) in the form of a free chat. The participant should not be stressed or pressured in any way since this is relaxed speaking practice with a native Japanese speaker that is supposed to help them improve in their conversation skills.

This system is assumed to be more engaging for the ghost user in contrast with

classical learning by books and is expected to immerse the learner to the Japanese daily life in the most effective way.

3.5. Interaction Flow

The application provides an immersive experience that integrates Virtual Reality, sound, as well as useful vocabulary items for a particular daily life scenario for visualizing, sharing and engaging with a tandem partner's environment through VR headset connected to a 360-degree camera for in situ language learning.

Ideally, the system should support the function to take notes while in the lesson, i.e. taking notes by hand while using a VR headset controller as a *wand* and later being able to access those notes. This is reflected to help memorising words better and, most importantly, to keep new vocabulary and useful expressions in one's long-term memory.

Before the data collection procedure, all participants who used the VR setup were given a consent form to sign. They were clearly informed that their participation in the experiment was completely confidential and that they could withdraw from the research at any time without consequence.

As a small "thank you" for participating, each participant receives an Amazon card worth JPY 1000.

3.6. Setup

The following components were used in the system:

- Virtual Reality headset (that can display the footage smoothly and uninterrupted as possible);
- 360-degree camera (with an excellent frames per second rate - fps) and a high resolution capable to broadcast in 4k;
- Open Broadcast System - OBS - to make it possible to conduct a real time lesson without delay;

- A high speed internet connection for the lesson to go as undisturbed as possible to ensure the voice chat to go smoothly.

3.6.1 Hardware

In particular, these are the items we used for the prototype and the user study:

- Meta Oculus Quest (first generation) Virtual Reality headset;
- GoPro MAX 360-degree camera capable to film in 360-degree in 4k resolution. One of the critical advantages for this model was its tested decent image stabilization, high quality audio with 6 microphones, and a lightweight waterproof design perfect for our in situ lesson;
- After thorough consideration, it was decided not to use OBS was for neither of the prototypes in order to keep the virtual environment and the content the same for every participant. Instead, pre-recorded 360-degree footage was used to produce an engaging video with both voiceover and ambient sound to make it identical to a real life lesson;

3.6.2 Content

In terms of content, several possible scenarios were considered, mostly focusing on real life situation like buying groceries, commuting by public transportation, ordering and eating local food, getting around the campus, visiting a historical site/specific district in Tokyo. Even taking a walk (see figure 3.3)with in a country which language you are learning is expected to be an engaging and motivational experience for learners to use Japanese outside of traditional classroom. In the end we have come to a decision to select two scenarios: taking a train in the world's largest city (for the pilot user study) and going to a sushi restaurant in Tokyo (for the final prototype).

Figure 3.3 "Active user" side of setup: the teacher GoPro 360-degree camera and a suitable place outside in Japan for in situ lesson.

Figure 3.4 "Ghost user" side of setup: the learner wears a VR Oculus headset while having 360-degree footage casted on the device all the way from Japan.

Chapter 4
Proof of Concept

4.1. Overview

This chapter is going to describe in detail how the user studies were conducted, which scenarios were used in particular and why, as well as physical setup and results of experiments. The proof of concept is going to use almost the same physical setup as it would for the real life field lesson, but due to limitations and possibility of inconsistent results, we have concluded it will be best to use pre-recorded 360-degree footage and immerse the learners in that environment.

4.1.1 Developing Scenarios

First, a few possible scenarios were planned. Lesson scenario development is crucial because we are relying on pre-recorded footage for the user study. Taking relatable conversation ideas, context and environment of Japanese language in consideration, essential daily conversation topics were identified:

- Getting around by public transport in Tokyo;
- Favorite Japanese food and ordering it;
- Shopping;
- Unique Japanese matsuri (festivals);
- Japan's unique neighborhoods and districts to experience culture through;
- Lodging related topics.

These trendy conversational topics were then narrowed down to public transportation together with Japanese food and later evaluated into a narrative. We wanted to place the learner into a condition close to real life ones, ensuring the freedom for the learner to speak outside of the strict scenario. Instead of placing the learner into a sort of a movie plot, rather distinct touchpoints for the free conversation to revolve around were built up.

Finally, according to scenario structure, a certain sequence of events were recorded in Tokyo, Japan. For the Transportation scenario, Buying a ticket - Getting on a platform - Studying the directions - Getting on a train - Finding the right exit. For the restaurant scenario, Kaiten Sushi or Sushi-go-round restaurant was selected as a restaurant venue, so the sequence followed Get off the station - Finding the entrance from - Reading the signs - Booking a seat - Ordering food from the menu - Learning fish kinds off the conveyor - Getting the bill - Paying by cash at the counter.

4.1.2 User Study

The study was conducted during the spring semester of the academic year 2022 during the months of April and May. Several test runs were completed before using the final prototype, mostly because we wanted to ensure the quality of the footage and most importantly how stable the shot will be. The current thesis explains two different scenarios, i.e in commute in a public transport in Japan (1) and sushi go round restaurant experience (2).

There were 21 participants total that took part in the user tests conducted for this research. There were two user studies in total:

1. To test Prototype 1 (in this paper called Pilot user test hereinafter). A total of 5 participants. Public transportation scenarios were used in the Pilot test.

2. To test Prototype 2 (in this paper called Final user test). This prototype used Japanese food scenarios. In the frames of testing the second prototype, a total of 16 participants were invited: 8 for proposed VRx360 setup and 8 extra participants that made the focus group (we call it Conventional learning group in this paper). Both VR group and Conventional focus groups

used the same Japanese restaurant narrative, grammar points and vocabulary, but used different learning styles. Again, the VRx360 focus group studied the conversational topic by the use of proposed system, whereas the Conventional study group had a traditional lesson tete-a-tete with the teacher behind the desk. In order to see how effective Virtual Reality method in contrast with classic desk/blackboard method, another focus group was recruited in proportion to VR focus group number of participants. A small vocabulary + expression test in Japanese was offered at the end of the lesson.

Participants selected for the user study were adults mostly in their mid-late twenties, studying at a Master level and coming from various academic backgrounds. They all have studied more than one foreign language in their life and are currently at the point where they are looking to improve their conversation skills in Japanese. They did take formal Japanese classes on several occasions.

Before conducting the user studies, our participants were tested to examine their general learning prefereneces.

4.2. VARK Questionnaire For Learning Preferences

VARK (Visual, Aural, Read/Write, and Kinesthetic sensory modalities) test suggests what learning preference a person tends to have. It can be a valuable piece of information when testing new learning systems in order to understand each individual's learning style and tendencies better.

According to the questionnaire, the tendency towards *Visual* strategies in Education uses symbolism and different formats, fonts and colors to emphasize important points. It does not include video and pictures that show real images and it is not Visual merely because it is shown on a screen. Then *Aural* strategy is for "information that is spoken or heard. Making statements and using questions are an important for those with this preference". Moving on, *Read/Write* preference seems rarely a big number, but people with this strategy preference put the emphasis is on words and it is arranged into lists. People with this learning style believe that "the meanings are within the words, and that people need to

be careful when using words". Finally, the *Kinesthetic* modality suggests leaning to emotions, senses, practical exercises, experiences, examples, case studies, test and error. For learners with a kinesthetic preference it is easier to learn in the context that could be "relevant and concrete".

4.3. Prototype 1: Data Analysis and Exploring Users' Needs

This is the pilot study that was conducted in order to identify and formulate users' needs and what a "ghost user" wants to see in an online Japanese lesson like this while talking to a native speaker. For this prototype it was decided to use pre-recorded 360-degree footage in the *Commuting in Tokyo* setting. 5 learners of Japanese were invited for the user study, all based in Japan and coming with an Intermediate level of Japanese already, figure 4.9.

	age	do you have previous experience with VR?	Japanese level	Listening ability	Reading Ability	Speaking Ability	Grammar Ability	Writing Ability
1	20~25	yes	N3	modest	good	modest	competent	modest
2	20~25	no	N3	extremely limited	competent	limited	modest	limited
3	26~30	yes	N2	good	very good	competent	good	very good
4	26~30	yes	N2	limited	modest	extremely limited	extremely limited	extremely limited
5	26~30	yes	N2	good	llimited	modest	intermittent	intermittent

Figure 4.1 Pilot user test: subjective user assessment of their Japanese language proficiency (total 5 participants).

Figure 4.2 User Test Setup: The ghost user is transported to Japan to learn Japanese at an online tandem lesson.

4.3.1 Purpose & Hypothesis

The purpose of this study is to get acquainted with the current learners' expectations and whether the design proves itself to be effective. The tutor was an adult female who is a native Japanese speaker and is currently conducting research at Keio University. The learners are mostly intermediate learners of Japanese that have been living in Japan for at least 1 year but, due to academic and work surrounding, they do not have much exposure to Japanese language. Participants were also asked to complete the learning preferences VARK test (visual/aural/read&write/kinesthetic) and you can see the distribution among the participants on figure 4.3.

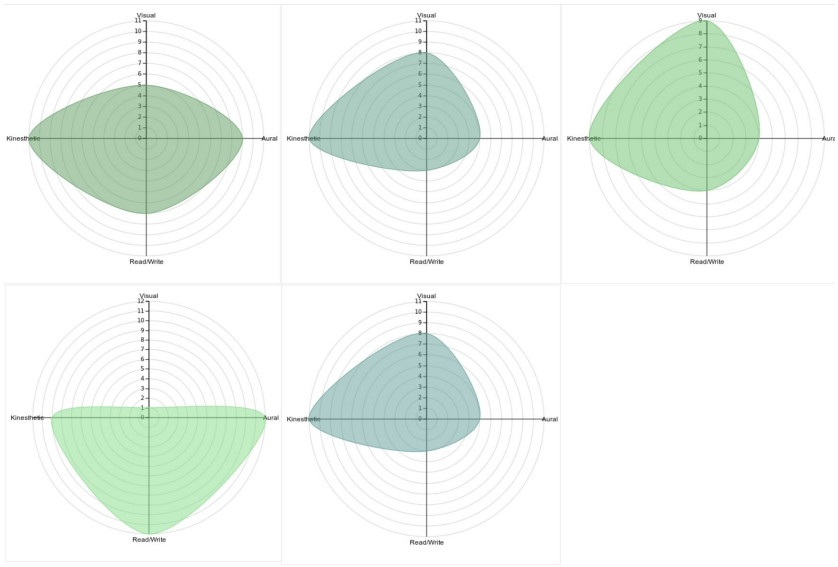

Figure 4.3 Pilot User Test: learning preferences based on visual, aural, read/write, and kinesthetic sensory modalities.

4.3.2 Design & Content

A pre-recorded Tokyo setting was recorded in a format of a 360-degree video (with original ambient sound) of about 20 min long was casted locally to a ghost user in a soundproof room via a VR headset. The video itself had a scenario of

commuting by local transport in Tokyo to meet up with a friend in Roppongi on a Sunday afternoon. The video consisted of about 10 sequential scenes that were shot statically from a tripod. The reason for filming from a tripod is that the previous footage (shot off a free hand) turned out to cause harsh motion sickness as the camera movements were not in sync with the body movements. While not people suffer from motion sickness, it was decided to do a static film with a downside of being less engaging. We hoped that engagement component in this setup could be compensated by the opportunity of a free conversation with a Japanese native speaker, but that was not the case (see results).

The participants were told to imagine that they are back in their country, sitting comfortably in the living room. They are participating in this online immersive Japanese learning lesson with a Japanese friend. Although no preparation was needed, the participants were told before the experiment that they are going to practice vocabulary relevant to commute and directions/locations in Japan. The script was as follows: *You are studying abroad in Tokyo, Japan. You had just recently arrived to the country and it is your 2nd week here. Your Japanese friend that also happens to be your new classmate kindly helps you to get around. Today you are going to see Roppongi hills and walk all the way from Midtown to Mita campus of Keio university. Figure out how to buy a ticket, navigate among the platforms and confusing local/express JR trains while using as much relevant vocabulary as possible.* Afterwards they were reassured that they are encouraged to interact with the native speaker of Japanese and ask any word they might not know in that language.

4.3.3 Procedure

The participants were asked to follow the video scenario as the narrative would lead them from home station to the destination (Roppongi station) while conversing with a Japanese friend (an assistant for the pilot study that was connected to the research participant by a voice chat (on Zoom). Remarkably, the experiment did not go according to the strict scenario and it was hard to generate topics for conversation, although the language assistant was familiar with the contents of the video. As a result, users had to skip forwards at an interval of 10 seconds in order to change the setting to the next one.

What are the advantages and disadvantages of the learning with the use of VR that you have just finished?	Did you experience any nausea or dizziness throughout the experiment?
• The advantage is creating a almost real context compared to the traditional learning style, and the disadvantage is that VR headset got me a bit dizzy and it's kind of heavy.	• Yes, a little bit after 10 mins.
• Advantage is that I feel better interacting with surrounding stuffs without feeling limited. The disadvantage is that the timing of some scenes is too long without real time interactions	• a little bit but it was ok
• Felt alone, not as nice as when there is a real person by me	• no
• Advantages: Immersive environment, more real interactions, super close to real world context. Disadvantages: eyes feel tired after a long time, No people interaction, sometimes boring if you stay in the same place for too long.	• yes, while moving
• Not very interactive, would like to see more hints on what to talk about.	• Yess, felt a bit. But I think 20min is okay.

Figure 4.4 Pilot user test: learner ' s free comments on the flow of experiment and idea of proposed system.

4.3.4 Results

Post-experiment results on perception, motivation and engagement were self-assessed by the 5 participants in total based on a Likert scale from 0 = strongly disagree to 5 = strongly agree.

Perception results see figure 4.5 represent participants' perceived abilities in holding a conversation in Japanese, in particular Grammar, Casual forms, Locations, Objects, Interior and Exterior, as well as Directions. Overall the results are lower than expected, that is reflected to be due to lower, than average engagement levels (see figure 4.7.

The scenario was quiet vague because we assumed the participant will want to use as much time as possible to interact with the Japanese speaker. The "ghost" user participants that we had tested it on said that they'd like to see even more prompts to spark a conversation and they also felt left alone in their adventure, apparently because fast content is preferred nowadays and train commutes are not as exciting anymore for those living in Japan – all of the participants are residents in Kanto area.

Perception assessment revealed how well the instructions were given and to what extent the participants are able to give/follow these topics after an immersive real time VR lesson. Overall, it is lower than expected. Out of 5 participants, motivation levels appear to be considerably higher than engagement results. However, self assessed results for motivation levels were significantly higher – see figure 4.6, that reflects a good potential and interest for studying in the Virtual Environment.

4.3.5 Post-Experiment Interviews

After each session an interview was conducted in a rather relaxed manner where we asked about general advantages and disadvantages of the current prototype. Although 80% of respondents pointed out a good potential for the system, over a half of participants mentioned it was not as engaging as it could be and also they found it too monotonous to stay in the same scene for too long. Despite the fact that every setting won't last longer than 2,5 min, and this time was tailored for any free conversation that might occur, the majority wanted to see something more dynamic.

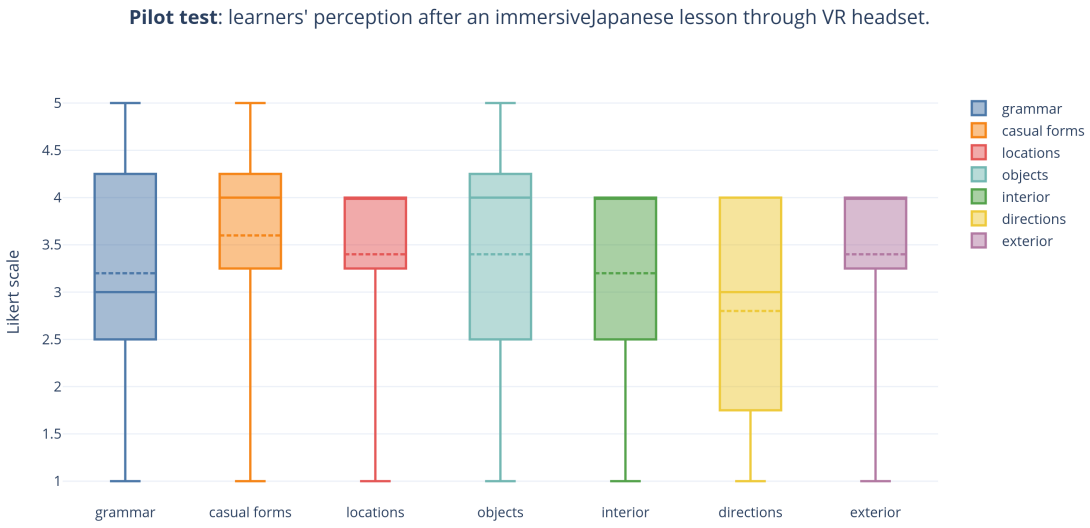

Figure 4.5 Pilot user test: learners' perception after having an immersed lesson through VR headset regarding learning Japanese with a tandem language partner, where participants were asked to assess their perceived confidence on using New grammar, Casual forms, Locations, Objects, Interior, Directions, Exterior. (from 0 = strongly disagree to 5 = strongly agree) — 5 participants in total.

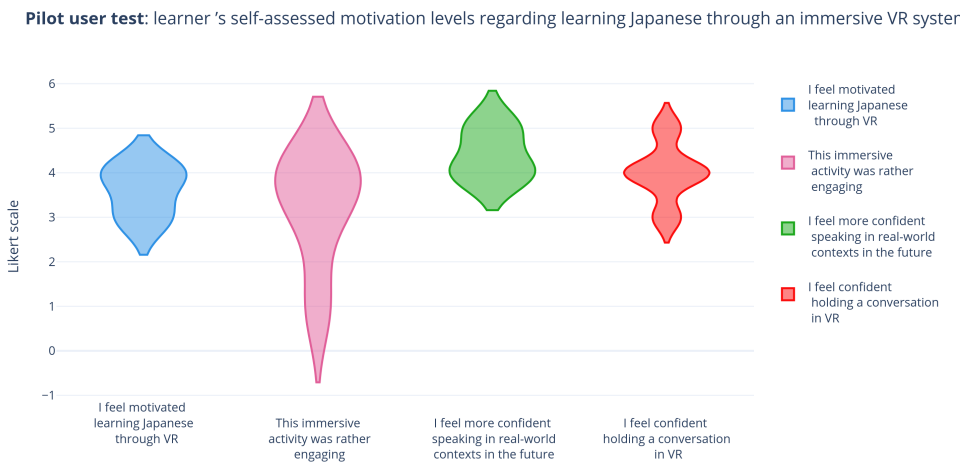

Figure 4.6 Pilot user test: learners' motivation levels after completing an immersive Japanese lesson through a VR headset where they were asked to assess how motivated they feel to (blue) Learn Japanese through VR, (pink) Engage to learn with VR, (green) Speak Japanese in the real world settings in the future and (red) Hold a conversation in VR (from 0 = strongly disagree to 5 = strongly agree) — 5 participants in total.

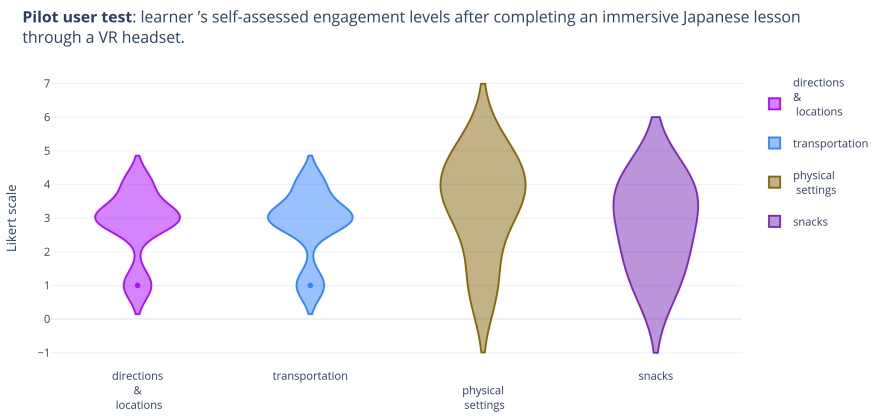

Figure 4.7 Pilot user test: learners' engagement levels after completing an immersive Japanese lesson through a VR headset (Directions Locations, Transportation, Physical settings, Snacks (from 0 = strongly disagree to 5 = strongly agree) — 5 participants in total.

4.3.6 Discussion

Keeping in mind the classical theory for contextualized learning by Vygotsky [16], the initial user study did have the environment, the objects, however it hardly had any individuals, and therefore interaction between people and above mentioned elements of one system was impossible. Feedback from the users only confirms that: *I felt like I am left alone in my learning adventure. Although I did have a person to support me over a phone call, I could clearly understand that it was not the same and so I did not feel motivated to learn in any way.*

The results are overall lower than expected and too much freedom of the conversation flow reflects the week point of proposed system as our learners admitted that they could not come up with the topic of conversation and even felt uncomfortable speaking to the tutor they have not interacted with before. More thorough preparation for the tutor to fill the awkward pauses might be a solution.

Thanks to this data, we got the valuable input that everyone's learning style is different and, what can be an incredible chance to have a conversation with a native speaker and practice Japanese, for others lack of hints felt too uncomfort-

able. Another reason why the results are not as impressive as they could be for the pilot study is the fact that every participant was tested in Japan, meaning they have already been here for some time and commuting on a train to casually meet a friend might not be as captivating as it could have been.

According to the feedback gathered in the session, we have identified the most attractive learning scenarios for Japanese learners that are already settled in Japan (that could be an uneasy task). It appears that the users favor something relatable and rather straightforward, and based on the input the following preferable scenarios are:

- Ramen shop: engaging enough yet about authentic food;
- Akihabara district: second hand electronic stores/figures/manga/collectables;
- Sushi-Go-Round restaurant: ordering process, matcha tea introduction;
- Station: the process of buying a ticket, putting the ticket in the machine, buying something at the station convenience store, studying the labels written on the platform (先発・後発);
- Busy intersection: studying the surroundings, reading the ads in kanji.

4.4. Final Prototype: Towards Engagement and Immersion

Having collected all the feedback during the pilot study, we have come to fundamentally different approach for the final prototype. In the era of fast content this always-on world, as well as the massive volumes of content available on the internet has serious repercussions for the modern consumer's attention span. The optimal time for the 360-degree video lesson experience is about 10 minutes but no longer than 15 minutes for one scenario.

Consequently, for the final user study 回転寿司 (kaiten zushi) restaurant, or a sushi-go-round restaurant was selected given its comprehensive display of the new words and expressions related to eating local traditional Japanese food. In addition, the conveyor belt system being entertaining on its own therefore this

Figure 4.8 Final user test: a screenshot from the user study - different types of fish in a sushi-go-round restaurant.

location cannot by any means be called boring. The target vocabulary is ordering, fish names, eating at a restaurant.

4.4.1 Purpose & Hypothesis

The purpose of this user test is to see how effective is the 360 x VR system in terms of Japanese learning in the context of Japan environment as well as studying the perception of VR x 360 custom made environment in the frames of Japanese language learning. In order to measure the results and have a piece of feasible data to play with, participants would be offered to take a short test consisting of 23 questions about the content of the video and new vocabulary. This data will be later on compared to a conventional study group that is going to cover the same curricula but in a traditional format.

4.4.2 Design

Aiming at the prototype to be close to real life settings as possible, the final one was upgraded to a thoroughly edited video that would be captivating to watch on its own. Furthermore, interactive voiceover by the clip creator was added in a way that it will spark a two-way conversation. Along with that, 3D short messages are added within the casted video containing the same vocabulary disscused orally but in a written form - see fig.

4.4.3 Content

While the design has stayed the same from the pilot study, the content of the final experiment was changed. As mentioned, we chose a newly opened Sushi-go-round restaurant (Kura sushi) in the heart of Harajuku. Thus we expected to also cover a little more of commute and expose the learners to a busy street of Harajuku while introducing the lesson's topic. The participants are now engaging with the person in video, meaning there is not behind the scenes language tandem partner that supports the conversation.

Figure 4.9 Final User Test: The ghost user is asking about the sushi types that the active user is showing them at an online Japanese lesson.

4.4.4 Procedure

The final user study was conducted during May-June 2022. After a brief orientation, the participants were asked to interact with VR environment and the Japanese language partner ("active user") through watching 360-degree video related to visiting a sushi restaurant in Japan. The list of vocabulary and expressions were not taught before and no prior preparation was needed for the test.

4.4.5 Results

In order to examine perception (see figure 4.10) of studying in the VR x 360 environment, Likert-style questionnaires were offered upon completion of immersive lesson. Participants were asked to assess their perceived confidence on using New grammar, Casual forms, Locations, Objects, Interior, Directions, Exterior. The mean number was considerably higher than the number from the Pilot study. Remarkably, the sushi restaurant perception related levels were on a higher end. Second, in terms of language acquisition and practice, participants stated that if they did not live in Japan currently, it would have been an invaluable tool for learning a foreign language.

Motivation related levels (fig. 4.11) appeared to be a bit higher than the perception related levels. Remarkably, sushi restaurant perception related levels were on a higher end, which proves the hypothesis: VR x 360 system can indeed engage the user into its environment and contribute best for learning the relevant to the context vocabulary as well as retaining it in the long-term memory. In addition to perception levels, engagement levels (see figure 4.12 confirmed learners' motivation levels with the average good above 4, especially for motivation on learning Japanese through VR, hold a conversation in Virtual Environment and speaking Japanese in the real world settings.

4.4.6 Discussion

The students have mostly perceived positively the ability of Oculus VR headset activities, despite the fact that it was physically heavy for some participants and generally the quality of displayed content is rather grainy given the headset is from the previous generation. Most participants strongly agreed (5 out of 5) or agreed

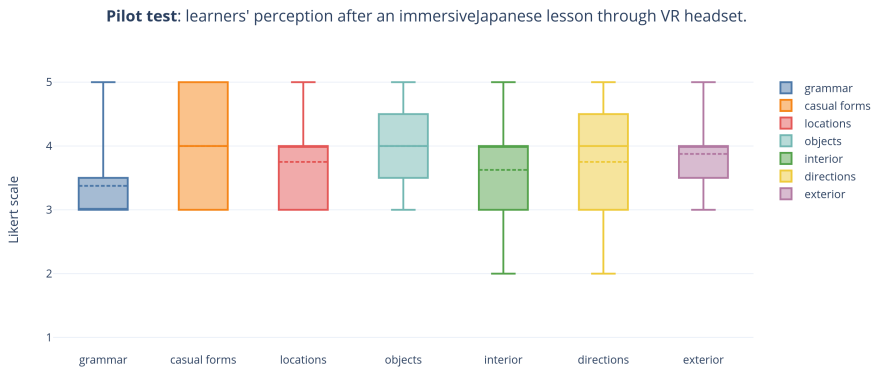

Figure 4.10 2. Final user test: learners' perception after having an immersed lesson through VR headset regarding learning Japanese with a tandem language partner, where participants were asked to assess their perceived confidence on using New grammar, Casual forms, Locations, Objects, Interior, Directions, Exterior. (from 0 = strongly disagree to 5 = strongly agree) — 8 participants in total.

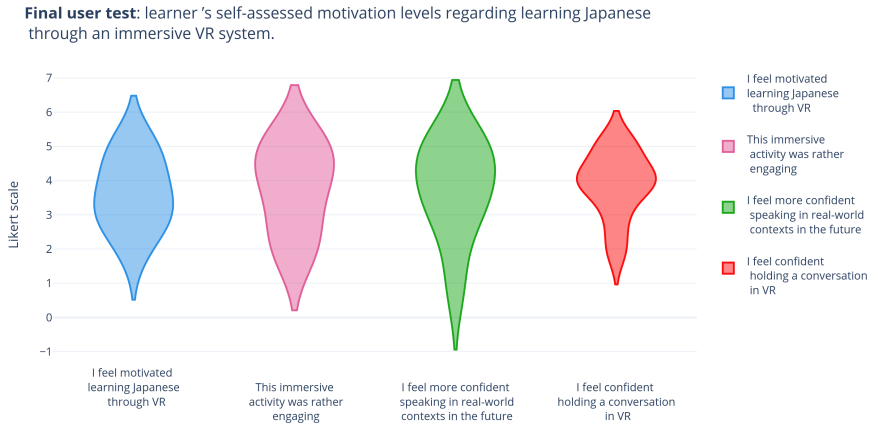

Figure 4.11 3. Final user test: learners' motivation levels after completing an immersive Japanese lesson through a VR headset where they were asked to assess how motivated they feel to (blue) Learn Japanese through VR, (pink) Engage to learn with VR, (green) Speak Japanese in the real world settings in the future and (red) Hold a conversation in VR (from 0 = strongly disagree to 5 = strongly agree) — 5 participants in total.

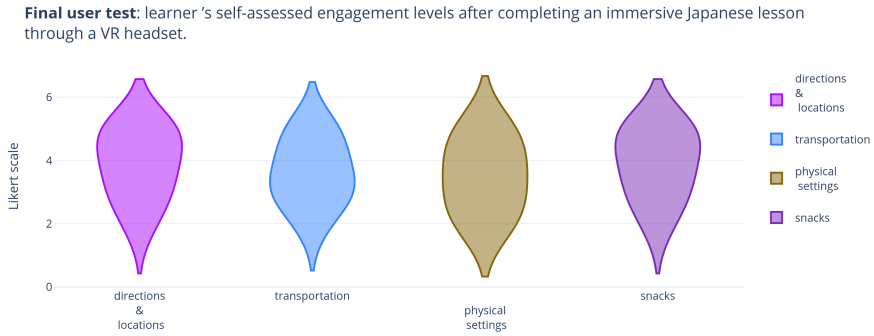

Figure 4.12 3. Final user test: learners' engagement levels after completing an immersive Japanese lesson through a VR headset (Directions Locations, Transportation, Physical settings, Snacks (from 0 = strongly disagree to 5 = strongly agree) — 8 participants in total.

(4 out of 5) that their language learning improved through the VR activities. The majority of participants highly agreed that they understood and could use specific targeted language and conversational Japanese.

Second, in terms of language acquisition and practice, participants stated that if they did not live in Japan currently, it would have been an invaluable tool for studying the most difficult skill while learning a language at home - speaking. One participant had mentioned that it was very helpful when it comes to learning short forms and applying them in every day Japanese (like 〜だよね ending) for which there's no textbook training. Overall,the majority of participants indicated that the authentic VR environments provided using 360 degree video from Japan made them feel as if they had been given the opportunity to apply what they had learned to real-world problems.

On a funny side, several participants had suggested that perhaps the angle of the camera was not well-thought since they felt like they were one of the sushi on a plate (how humongous the plates seemed through a VR headset).

However, in order to be able to validate the current positive results, it is crucial to conduct another study focus group that would try to learn the same target vocabulary but in a traditional setting.

4.4.7 Post-Experiment Interviews

Finally, after the lesson and all the standardized questionnaires we asked for the general feedback, including questions like what participants think potentially be improved, what did they fancy or not throughout the experiment or any free comments or questions they might have to us.

Some participants have also reported that the font was too big for them while they were wearing VR headset. In addition to that, it was not always 100% clear which direction to look at to see the floating text, so that could be a point to improve as well.

4.5. Focus Group: Conventional Learning Style

The traditional teaching of foreign languages still relies on textbooks and rarely unites with technology, other than for video conferencing. As mentioned earlier,

in order to get a fair impression on how students comprehend new vocabulary material, a separate focus group is needed in order to evaluate how well the 350 x VR method works. To make the results evident, we decided to deploy a small vocabulary + grammar expressions test after the study. We then compare results of both focus groups and discuss what are the possible take aways are.

4.5.1 Design

The lesson was designed in such way that it would include all traditional techniques that are commonly accepted and utilised globally in the classroom. Essentially, the lesson had a set of slides on a particular topic and students could take this lesson both individually and in small groups.

4.5.2 Content

In order to be on the same page with the Kaiten Zushi experiment, this lesson was tailored also based on food and primarily focused on naming different Japanese foods and relevant expressions.

Figure 4.13 Conventional focus group: materials from the lesson presentation.

4.5.3 Procedure

A total of 8 participants were offered a classical lesson (i.e. learning from a teacher and with the help of a textbook and other writing materials) in Japanese with the content identical to the immersive user study. The participants were informed before the lesson that they are going to be tested at the end of the session and that it is in their best interest to memorize as much as they can from this lesson. It is also necessary to mention, that the participating students were encouraged to interact with the teacher and ask them questions all to see the best results they are capable of delivering.

4.5.4 Results

After correlating results of the small test (figure 4.14 and 4.15), it is evident that the suggested novel VR x 360 method for learning specifically in context and while being immersed in country's local neighborhood, that all can play a deciding role in effectiveness of learning foreign languages. Participants, that have completed the VR immersive session had scored almost twice as better against those that did the conventional study. It is anticipated that even if the teacher is using real photos on their presentation, it might not be enough to stick that to students' memory despite the repetitions.

question	max points
Write down as many new types of sushi as you could possibly remember from the lesson. Only those that were introduced/mentioned will count.	5
How to pronounce "sushi go round" restaurant correctly in Japanese?	1
What does it mean インスタ映え?	1
Tuna in Japanese:	1
Salmon in Japanese:	1
Shrimps in Japanese:	1
Sweet shrimps in Japanese:	1
Omelette sushi in Japanese:	1
"Time to pay" in Japanese:	2
Minced tuna in Japanese:	1
What does it mean, 回ってくる?	2
Green tea in Japanese:	1
How to say "do you recall" in Japanese?	2
All-you-can-eat in Japanese:	1
All-you-can-drink in Japanese:	1
What is 赤貝（あかがい）in Japanese?	1
	23

Figure 4.14 A list of questions from the small post-experiment test that were used to correlate the effectiveness of VR based in situ learning over conventional study methods.

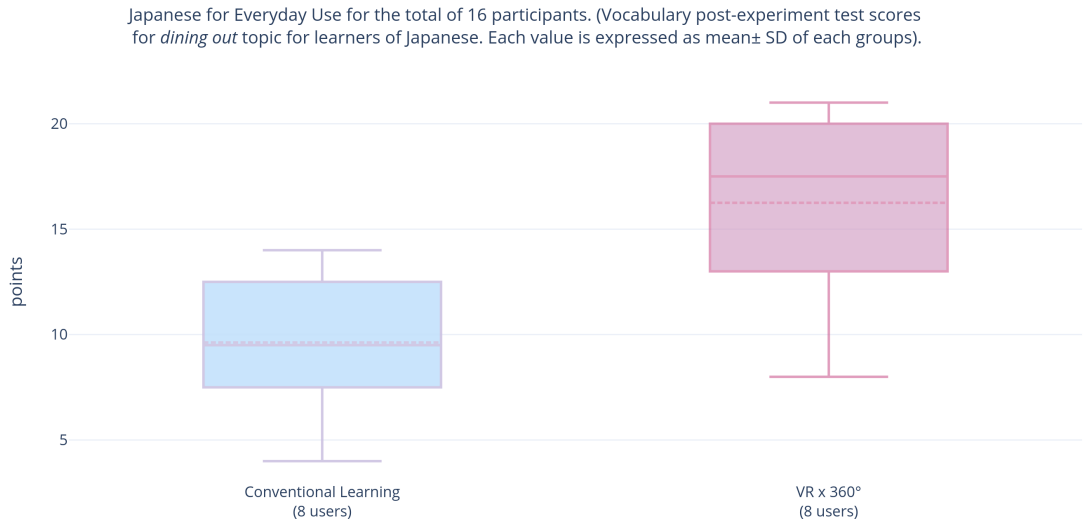

Figure 4.15 Conventional focus group: materials from the lesson presentation.

Chapter 5
Conclusion

5.1. Summary

This thesis offers a system that connects two language tandem partners by broadcasting 360 degree footage shot in real-life by the tutor ("active user") to a VR headset of a learner ("ghost user"), thus providing an exceptional immersive language experience without leaving the house. The method is proven to be effective in terms of improving one-on-one conversation practice, which, in turn, will help to retain more vocabulary items and most importantly to extract that piece of knowledge from the depths of the memory next time the learner will find themselves in a similar situation. The current prototype can be already considered as additional help in learning Japanese.

Although Virtual Technology is being widely used nowadays and it in fact becoming more affordable, there are still very few studies that use VR in regards with language learning, and if they do, it is rather a small percentage focuses on language learning (mostly English as a second language and Chinese) [11] [38]. Besides, that existing research is tested on a smaller focus group and is being in an embryonic stage as a research direction. However, VR has a huge potential to transform how we learn foreign languages in the near future.

We have tested different scenarios and, while it highly depends on each individual's preference, widely known activities that anyone can relate to seem to be most popular regardless of their age, sex, country of origin. Those are, to name a few, eating out, using public transportation, going shopping,- simple and motivating to learn more at every level of language proficiency. We have also found out that the pace is very important nowadays in the era of fast content - people's attention span is very low, so even while wearing such an immersive device, it is recommended for the tutor ("active user") to still be more encouraging to speak

up as well as changing locations quickly.

The voices of our participants have confirmed the effectiveness of the current system as well as a high degree of involvement in the situation. Because all the participants daily use video calling platforms such as Zoom, Skype, Line, etc., they naturally noticed and pointed out on the differences in interactions between VR x 360 and 2D screen teleconference. The main take away is that immersive content in the real world can enhance the learning compared to a VR environment that has nothing to do with the topic of conversation. That is because of the context - context plays a crucial role as we managed to reach interaction people within the cultural or historical setting while developing learners' cognition.

Finally, despite the proven benefits of proposed system, implementation still needs improvement, specifically on the technical setup side. The lighter the VR headset is, the more seamless it is reflected to be for the learner and therefore will cause less distractions (supporting the headset with a hand) during the learning process. Additionally, poor graphics performance on the first generation Oculus headset is reflected to distort original footage and besides the blurry image, it causes motion sickness, which disorients the learner in the environment and does not correlate what they see with the body movements, therefore the brain cannot make sense of information sent through the headset with the static body position. Camera stabilization is still needs enhancing.

5.2. Limitations

It is important to point out that these results cannot be extrapolated due to a rather small group of participants. In addition, an issue we had faced during the experiment is that Oculus headset is heavy and for some people it is impossible to enjoy the lesson because they had to hold it against the head. Many have noticed that the headset was too heavy as well.

The main limitation, however, is regarding the dizziness and nauseous feeling participants get when they watch 360-degree video for longer than 10 minutes. That is a short period of time, yet impossible for some to survive.

5.3. Future Work

Immersive 360-degree camera supported VR based learning shows great potential in education and especially in foreign language learning. However, little actual research seems to exist today. For the future work, the actual set up and experiment itself is yet to be thoroughly planned and carried out. Evaluation of the current method is yet to be finalized.

It would be curious to conduct this study in slightly different settings:

- While the pointers are already working in the system, incorporating a writing tool (similar to a magic wand) while in the lesson could be another convinient add-on. This way the learners could make small notes at the side (new vocab) and later on accessing those notes after the lesson. This is reflected to be something that would elevate the control test scores even higher against conventional study.

- Applying this methodology outside of the country's language (in case of Japanese - the learner should be outside of Japan to fully immerse in the environment solely through VR.

References

[1] M. D.; Furst E. J.; Hill W. H.; Krathwohl D. R. Bloom, B. S.; Engelhart. *Taxonomy of educational objectives: The classification of educational goals. Vol. Handbook I: Cognitive domain. New York: David McKay Company.*, volume Vol.1 Handbook 1: Cognitive domain. New York: David McKay Company., 1956.

[2] Mondly. Play your way to learn a new language. URL: `https://www.mondly.com/`.

[3] Aleshia Taylor Hayes, Tetyana Kucher Dhimolea, Nanxi Meng, and Geneva Tesh. *Levels of Immersion for Language Learning from 2D to Highly Immersive Interactive VR*, pages 71–89. Springer Singapore, Singapore, 2021. URL: `https://doi.org/10.1007/978-981-16-3416-1_4`, `doi:10.1007/978-981-16-3416-1_4`.

[4] Microsoft Corporation. What is mixed reality? URL: `https://docs.microsoft.com/en-us/windows/mixed-reality/discover/mixed-reality`.

[5] Abeer Madini and Dalal Alshaikhi. Virtual reality for teaching esp vocabulary: A myth or a possibility. *International Journal of English Language Education*, 5:111, 10 2017. `doi:10.5296/ijele.v5i2.11993`.

[6] Regina Kaplan-Rakowski and Alice Gruber. *One-On-One Foreign Language Speaking Practice in High-Immersion Virtual Reality*, pages 187–202. Springer Singapore, Singapore, 2021. URL: `https://doi.org/10.1007/978-981-16-3416-1_9`, `doi:10.1007/978-981-16-3416-1_9`.

[7] Empsen M. & Hustinx W. A Evens, M. Systematic literature review on 360-degree video as an educational tool: towards design guidelines. *J.Computer*

Education, 18, 2022. doi:10.1007/s40692-022-00233-z.

[8] Tilman Dingler, Dominik Weber, Martin Pielot, Jennifer Cooper, Chung-Cheng Chang, and Niels Henze. Language learning on-the-go: Opportune moments and design of mobile microlearning sessions. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '17, New York, NY, USA, 2017. Association for Computing Machinery. URL: https://doi.org/10.1145/3098279.3098565, doi:10.1145/3098279.3098565.

[9] U.S. Department of State. Fsi ' s experience with language learning. URL: https://www.state.gov/foreign-language-training/.

[10] William E. Nagy, Patricia A. Herman, and Richard C. Anderson. Learning words from context. *Reading Research Quarterly*, 20(2):233–253, 1985. URL: http://www.jstor.org/stable/747758.

[11] Jaziar Radianti, Tim A. Majchrzak, Jennifer Fromm, and Isabell Wohlgenannt. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147:103778, 2020. URL: https://www.sciencedirect.com/science/article/pii/S0360131519303276, doi:https://doi.org/10.1016/j.compedu.2019.103778.

[12] Miao-fen Tseng and Ziyi Geng. *Implementing Virtual Reality–Enhanced Tasks in Chinese Language Teaching*, pages 91–118. Springer Singapore, Singapore, 2021. URL: https://doi.org/10.1007/978-981-16-3416-1_5, doi:10.1007/978-981-16-3416-1_5.

[13] İlknur Reisoğlu, F. Topu, Rabia Yilmaz, Turkan Karakus Yılmaz, and Yuksel Goktas. 3d virtual learning environments in education: a meta-review. *Asia Pacific Education Review*, 18, 02 2017. doi:10.1007/s12564-016-9467-0.

[14] Ajay Bharathi and Conrad Tucker. Investigating the impact of interactive immersive virtual reality environments in enhancing task performance in online engineering design activities. page V003T04A004, 08 2015. doi:10.1115/DETC2015-47388.

[15] Murat Akçayır and Gökçe Akçayır. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review*, 20:1–11, 2017. URL: https://www.sciencedirect.com/science/article/pii/S1747938X16300616, doi:https://doi.org/10.1016/j.edurev.2016.11.002.

[16] LS Vygotsky. Istoriya razvitiya vysshikh psikhicheskikh funktsiy [the history of the development of higher mental functions]. *Sobr. soch*, 6, 1983.

[17] Sigurdur O. Adalgeirsson and Cynthia Breazeal. Mebot: A robotic platform for socially embodied presence. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*, HRI '10, page 15–22. IEEE Press, 2010.

[18] Susumu Tachi. *Telexistence*, pages 229–259. Springer International Publishing, Cham, 2015. URL: https://doi.org/10.1007/978-3-319-17043-5_13, doi:10.1007/978-3-319-17043-5_13.

[19] S. Tachi, K. Minamizawa, M. Furukawa, and C. L. Fernando. Telexistence — from 1980 to 2012. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5440–5441, 2012. doi:10.1109/IROS.2012.6386296.

[20] Susumu Tachi. From 3d to vr and further to telexistence. In *2013 23rd International Conference on Artificial Reality and Telexistence (ICAT)*, pages 1–10, 2013. doi:10.1109/ICAT.2013.6728898.

[21] Minori Manabe, Daisuke Uriu, Takeshi Funatsu, Atsushi Izumihara, Takeru Yazaki, I-Hsin Chen, Yi-Ya Liao, Kang-Yi Liu, Ju-Chun Ko, Zendai Kashino, Atsushi Hiyama, and Masahiko Inami. Exploring in the city with your personal guide:design and user study of t-leap, a telepresence system. In *19th International Conference on Mobile and Ubiquitous Multimedia*, MUM '20, page 96–106, New York, NY, USA, 2020. Association for Computing Machinery. URL: https://doi.org/10.1145/3428361.3428382, doi:10.1145/3428361.3428382.

[22] Sven Kratz, Don Kimber, Weiqing Su, Gwen Gordon, and Don Severns. Polly: "being there" through the parrot and a guide. In *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices amp; Services*, MobileHCI '14, page 625–630, New York, NY, USA, 2014. Association for Computing Machinery. URL: https://doi.org/10.1145/2628363.2628430, doi:10.1145/2628363.2628430.

[23] Tadakazu Kashiwabara, Hirotaka Osawa, Kazuhiko Shinozawa, and Michita Imai. Teroos: A wearable avatar to enhance joint activities. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, page 2001–2004, New York, NY, USA, 2012. Association for Computing Machinery. URL: https://doi.org/10.1145/2207676.2208345, doi:10.1145/2207676.2208345.

[24] Yuichi Tsumaki, Fumiaki Ono, and Taisuke Tsukuda. The 20-dof miniature humanoid mh-2: A wearable communication system. In *2012 IEEE International Conference on Robotics and Automation*, pages 3930–3935, 2012. doi:10.1109/ICRA.2012.6224810.

[25] Luis Almeida, Paulo Menezes, and Jorge Dias. Telepresence social robotics towards co-presence: A review. *Applied Sciences*, 12(11), 2022. URL: https://www.mdpi.com/2076-3417/12/11/5557, doi:10.3390/app12115557.

[26] Shunichi Kasahara, Shohei Nagai, and Jun Rekimoto. Jackin head: Immersive visual telepresence system with omnidirectional wearable camera. *IEEE Transactions on Visualization and Computer Graphics*, 23(3):1222–1234, 2017. doi:10.1109/TVCG.2016.2642947.

[27] Atsushi Otaki Kosuke Takahashi Kentaro Itonori Yusuke Kamiyama Hazuki Sugai Takushi Okina Masae Suno Tomohiro Sawada, Ory Yoshifuji. Nin nin: Body sharing robot. URL: https://bodysharing.net/#project.

[28] MHD Yamen Saraiji, Tomoya Sasaki, Reo Matsumura, Kouta Minamizawa, and Masahiko Inami. Fusion: Full body surrogacy for collaborative communication. In *ACM SIGGRAPH 2018 Emerging Technologies*, SIGGRAPH '18, New York, NY, USA, 2018. Association for Computing Machin-

ery. URL: https://doi.org/10.1145/3214907.3214912, doi:10.1145/3214907.3214912.

[29] Chen Yu-Li. The effects of virtual reality learning environment on student cognitive and linguistic development. In *Asia-Pacific Edu*, Res 25, pages 637–646, 2016.

[30] Deborah L Ulrich Eric Hodgson Stephanie Nicely Angelia Mickle Sherrill J Smith 1, Sharon L Farra. Effectiveness of two varying levels of virtual reality simulation. *Nursing education perspectives*, 39(6):E10–E15, 2018.

[31] Zahira Merchant, Ernest T. Goetz, Lauren Cifuentes, Wendy Keeney-Kennicutt, and Trina J. Davis. Effectiveness of virtual reality-based instruction on students' learning outcomes in k-12 and higher education: A meta-analysis. *Computers & Education*, 70:29–40, 2014. URL: https://www.sciencedirect.com/science/article/pii/S0360131513002108, doi:https://doi.org/10.1016/j.compedu.2013.07.033.

[32] Emma Marsden and Carole Torgerson. Single group, pre-post test research designs: Some methodological concerns. *Oxford Review of Education*, 38:583–616, 10 2012. doi:10.2307/41702779.

[33] Margot A. Lazow, Francis J. Real, Nicholas J. Ollberding, David Davis, Bradley Cruse, and Melissa D. Klein. Modernizing training on social determinants of health: A virtual neighborhood tour is noninferior to an in-person experience. *Academic Pediatrics*, 18(6):720–722, 2018. URL: https://www.sciencedirect.com/science/article/pii/S1876285918301700, doi:https://doi.org/10.1016/j.acap.2018.04.007.

[34] Seung Hwan (Mark) Lee, Ksenia Sergueeva, Mathew Catangui, and Maria Kandaurova. Assessing google cardboard virtual reality as a content delivery system in business classrooms. *Journal of Education for Business*, 92(4):153–160, 2017. URL: https://doi.org/10.1080/08832323.2017.1308308, arXiv:https://doi.org/10.1080/08832323.2017.1308308, doi:10.1080/08832323.2017.1308308.

[35] Danijela Barić, Grigore M. Havârneanu, and Cornelia Măirean. Attitudes of learner drivers toward safety at level crossings: Do they change after a 360° video-based educational intervention? *Transportation Research Part F: Traffic Psychology and Behaviour*, 69:335–348, 2020. URL: https://www.sciencedirect.com/science/article/pii/S1369847819301093, doi:https://doi.org/10.1016/j.trf.2020.01.018.

[36] Mildred Lopez, José Gerardo Carrillo Arriaga, Juan Pablo Nigenda Álvarez, Ricardo Treviño González, José Antonio Elizondo-Leal, Jorge E. Valdez-García, and Belinda Carrión. Virtual reality vs traditional education: Is there any advantage in human neuroanatomy teaching? *Computers & Electrical Engineering*, 93:107282, 2021. URL: https://www.sciencedirect.com/science/article/pii/S0045790621002640, doi:https://doi.org/10.1016/j.compeleceng.2021.107282.

[37] Karey Lan, Indy Hsiao, and Mei-Feng Shih. Effective learning design of game-based 3d virtual language learning environments for special education students. *Educational Technology & Society*, 21, 01 2018.

[38] Chris A. Boulton, Carmel Kent, and Hywel T.P. Williams. Virtual learning environment engagement and learning outcomes at a ‘bricks-and-mortar’ university. *Computers Education*, 126:129–142, 2018. URL: https://www.sciencedirect.com/science/article/pii/S036013151830174X, doi:https://doi.org/10.1016/j.compedu.2018.06.031.

Appendices

A. Consent form

CONSENT FORM

TANDEM LANGUAGE LEARNING USING 360 VIDEO USER STUDY

〒223-8526 横浜市港北区日吉 4-1-1 慶應義塾大学 | kitsing@keio.jp | Inga Kitsing

Dear Participant,

We are conducting a study to observe how does immersive virtual experience created by 360 degree video and VR technology foster language learning and to what extent. We are also curious about your motivation as a learner, engagement and progress as a learner of Japanese as a foreign language.

The study will help us to better understand the specific ways VR affects communication and situated learning and will, in turn, allow educators to create better virtual reality systems and environments using real environment footage.

If you agree to take part, in this study you will be immersed in a pre-recorded Japan environment and asked to communicate on a specific topic with a native speaker of Japanese. You are encouraged to use Japanese as much as you can and especially ask for new words and grammar structures to your tandem partner. You will then be asked questions about the content and your impression.

You will not be identified by name in any reports of the completed study. Any personal information obtained in connection with this study will remain confidential.
You can choose whether to be in this study or not. If you do volunteer, you may withdraw at any time. If you choose to withdraw from the study either during or immediately following your participation, your data will be destroyed and not included in the study.

I have read the consent form, received enough information regarding the study and agree to take part in this study.

NAME

DATE