

Language Learning through Multimodal Communication in VEC3D

Ya-Chun Shih & Mau-Tsuen Yang

Department of English, National Hualien University of Education, Taiwan;

Department of CSIE, National Dong Hwa University, Taiwan

Abstract

The aim of this research is to achieve multimodal communication in Collaborative Virtual Environments (CVEs) available for EFL learners' communicative competence development via text and voice chat, lip motion and body language. We assume that developing strategic competence should take precedence over the other communicative competence components. Lately, the proposed CVE with multimodal communication, named VEC3D, has supported the development of EFL learner strategic competence, the ability to use verbal and nonverbal communication strategies (CSs) to overcome communication breakdowns. A concept model for the future description and analysis of communication forms and CSs is proposed. Calling nonverbal CSs into interaction is essential for communication success. Nonverbal communication forms are also examined in the virtual communication context.

Keywords: Collaborative Virtual Environment, Multimodal Communication, EFL, VEC3D

Introduction

In our current research we are exploring how 3D Virtual Reality technologies might be used for the development of communicative competence in English in a college setting. With the advent of 3D Virtual Reality together with on-line 3D chat, computer vision and graphics technologies, our new generation 3D CVE named 3D Virtual English Classroom (VEC3D) enhanced the user's immersive experience via real-time written and voice-to-voice communications. We are primarily involved in pedagogical research and investigation surrounding the integration of these technologies to support English language learning and teaching.

Our attempt to ease and naturalize human communication over the Internet is

through multimodal - linguistic, actional and visual -communication in 3D virtual reality. We believe that interpersonal multimodal communication is critical for language learning through multimodal interaction integrating visual, aural, linguistic and nonverbal communication. With the help of computer vision and graphics technologies, including foreground and background segmentation, we intend to provide a multimodal environment with sophisticated and natural communications for EFL learners. VEC3D allows users to use the body language that is essential in natural communication and the slide shows commonly used in real lectures. There are two ways for users to use body language to convey meanings, either through real time audio or video or via an avatar.

To process sophisticated communication, a lecture video is captured in a typical classroom. The instructor is segmented from the background as a ROI (Region of Interest). The segmented ROI is then combined with a virtual classroom. Because network traffic is reduced by only transmitting the moving foreground in a video, the system can survive in a network with limited bandwidth. In response to language learning and use, VEC3D intends to offer additional options for communication in real time via voice and/or text with enhanced user freedom in displaying his or her gestures in processing sophisticated communication through paralinguistic information such as body posture. This research explores the pedagogical innovation and implication of teaching English as a foreign language via 3D VR.

Multimodal Perspective on Language Learning

Language learning is multimodal, arising from the interaction of visual, aural, linguistic and nonverbal communication. We created a multimodal user interface that allows the instructor to communicate with students using natural modalities, including gestures, facial expressions and lip motion. The new 3D Virtual Environment (VE) is synthesized as a connection of key technologies: Internet, 3D chatting, and graphical embodiment of users (Avatars). (See Figure 1.) The newest trends in 3D VE are more attainable and affordable to learners scattered across the country than before. Based on these new trends, we produced virtual language learning designs and environments for learners and educators using 3D VR technology.

The project is based on the theoretical assumptions underlying the current research in engaging learners and placing language learning in a meaningful context within which authentic learning experiences and real communication and interaction occur. Teachers of English to Speakers of Other Languages (TESOL) indicates, "Language is learned most effectively when it used in significant and meaningful situations as learners interact with others to accomplish their purposes. Language

acquisition takes place as learners engage in activities of a social nature with opportunities to practice language forms for variety of communicative purposes” (TESOL, 1997, p. 7).

Furthermore, multimodal interaction: text messaging, voice communication, lip contour extraction and gesture are assumed to effectively improve learners’ interaction and communication efficiency when communication occurs over the Internet. The whole communication process consists of the interplay between two modes: verbal and nonverbal communication. (Allen, 1999) Usually, communication difficulties arise because people receive relatively limited aural or visual sensory input. Speech is well-perceived through audio-visual integration and multimodal. (Massaro, 1998; Massaro & Stork, 1998; Chen & Rao, 1998) Both verbal and nonverbal CSs help language learners compensate for communication breakdowns. (Canale, 1983; Swain, 1984; Canale & Swain, 1980) With the emergence of advanced 3D VR, tracking and animated technologies, we witnessed natural human communication and sophisticated language use in 3D virtual multimodal environment.

Based on constructivist perspectives, our learners derived great benefit from 3D exploring and chatting in a community. The development of communicative competence is believed to be situated in interaction, communication and challenging goal-based activities. The way of achieving successful second or foreign language acquisition depends on whether a learner receives appropriate communicative input in accordance with his or her level of linguistic competence. (Krashen, 1985) A fluent native English speaking instructor joined in the project and helped students receive communicative input, facilitating proper English usage via multimodal communication channels.



Figure 1. Video Conferencing and Avatars to Complement the Lack of Nonverbal Clues in the CVE

Educational and Technical Importance of the Study

This project is important in multimodal communication inspiring EFL learners to take part in CVEs. This conceptually novel platform provides the potential to various settings, including language learning in classroom and commercial settings. CVEs facilitate prolonged immersion experiences, and extensive interaction highly required for writing and conversation class. Since nonverbal communication is critical to language teaching. In classroom settings, instructors frequently rely on nonverbals to convey messages, deliver knowledge, correct errors, manage a classroom, and offer extra contextual information associated with words. Because the multimodal CVE empowers language teachers through the provision of multimodal interaction integrating visual, aural, linguistic and nonverbal communication, it has become an innovative tool in language teaching.

VEC3D is constructed to maximize flexibility and responsiveness to evolving research and educational purposes. Based on our previous needs assessment and experiences results, new features were developed to generate and stimulate multimodal communication via text/voice chat, lip motion and body language in virtual reality. VEC3D allows users to show their body language, slides, text, video, audio, and virtual objects with the help of the state-of-the-art computer vision and graphics technologies. (See Figure 2 & 3.) In a teaching scenario with a live lecture, a camcorder captures the lecture video in a real classroom using a real instructor. The captured video is analyzed using computer vision techniques to distinguish the foreground (instructor) from the background (classroom) in real-time. The segmented foreground is then transmitted to remote clients and combined with the virtual classroom in VEC3D. Figure 2 shows an online lecture with a live instructor and his body language in VEC3D. Similarly, for a learning scenario with a group discussion, the actual face of each remote participant is captured by a webcam. Computer vision techniques discard the background, transmit the foreground (face) to other clients, and combine with the virtual meeting room in VEC3D. Figure 3 shows an online group meeting with three participants with live facial images in VEC3D. An innovative combination of multiple communication channels facilitates natural and social interaction, flexible teaching and learning and responsiveness to the needs of EFL learners and teachers.

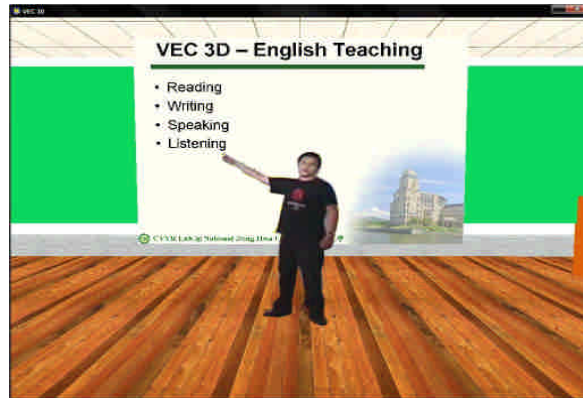


Figure 2. An instructor is experimenting with the use of multiple communication channels (slide, gesture, voice and text) offered by the multimodal CVE



Figure 3. Group discussion through Audio/Video conferencing

Multimodal Communication Tasks Embedded in the Virtual Environments

VEC3D is a foreign language learning and teaching platform that uses web-based 3D interactive multi-user environment to build virtual connection between a native English speaking instructor and students. The primary purpose of this project is to attract students through offering tasks that enhance their communicative competence. VEC3D project requires that students participate in both real-world and virtual activities, such as navigation, on-line discussion, team work and constructing virtual environment. Concurrently, we also constructed VEC3D in Activeworlds and used Second Life as a gateway to teleport to many other virtual 3D online communities, and meet other target language speakers around the world.

The VEC3D instructional design principles and theories are based on Constructivist Learning Design (CLD), Collaborative Learning (CL), Communicative Language Teaching (CLT), Goal-Based Scenarios (GBS) (Schank, 1992),

Story-Centered Curriculum (SCC) (Schank, 2003), Strategic Interaction (SI) (Di Pietro, 1987), simulation/game design for language learning (Crookall & Oxford, 1990), and virtual community building (Renninger & Shumar, 2002).

The tasks have been expanded for multimodal communication tools in hopes of stimulating nonverbal interaction. In addition, the new tasks include the activities for verbal communication, target cultural understanding, nonverbal expressions (body language, sound, and so on), and pronunciation practice (lip motion). All tasks were carried out within 60-120 minute class period. We intend to construct activities to engage EFL learners and build a virtual community of practice for communicative competence development. The following specific types of activities-- Icebreaker/Group Formation, Role-Playing, Exploration, Visual Activity, Jigsaw/Information-Gap, Drawing, Game/Simulation, Cross-Cultural Understanding, Action Game/Body Language, Pronunciation, Lecture, Group Discussion, and Virtual Field Trips-- are provided with several examples, forms of expressions (verbal or nonverbal), and communication channels (media) used. (See Table 1.)

Table 1. Summary of Online Activities in the CVEs

Types of Activities	Examples: (Mission)	Expressions (Main)	Media
Icebreaker/ Group Formation	Don't say that!	Verbal;	Text /Voice Chat
	True or False	Nonverbal: physical	Avatar(-Masking)
	Greeting & Interview Jason	appearance,	
	Guess who it is	kinesics-gestures, paralanguage	
Role-Playing	Short-Term Travel Study	do.	Text / Voice Chat
	"Dear John"		Avatar
	Job hunting		
	Pull-Out instruction		
	Complain about Food		
Exploration	Hide & Seek	do.	Text /Voice Chat
	Exercise & Explore		Avatar
	Scavenger Hunt		
	Exploring VEC3D		
Visual Activity	He isn't telling the truth	do.	Text/ Voice Chat
	What's behind you?		Avatar
	That's an unusual view!		Slides
	What's difference?		Pictures
Jigsaw/	Guessing Password	do.	Text /Voice Chat

Information-Gap	Odd man out		Avatar, Drawing
Drawing	Drawing and Imagination	do.	Voice Whiteboard
Game/Simulation	Rescuing Miss Saigon	do.	Text /Voice Chat
	Virtual Field Trips		Avatar
Cross-Cultural	Greeting & Interview Jason	do.	Voice Chat, Video
Action Game/ Body Language	Read your mind/lips	Verbal;	Text/ Voice Chat
	Distraction	Nonverbal: physical	Avatar, Video
	What on earth is Jason talking about?	appearance,	
	Vocabulary Pantomime	kinesics-gestures,	
	Simon says	body movement,	
Exercise & Explore	paralanguage, facial expressions-lip, mouth shape		
Pronunciation	Minimal Pairs	Verbal; Nonverbal: facial expressions-lip, mouth shape	Voice Avatar Lip Motion
Lecture	PPT presentation	Verbal; Nonverbal:	Text /Voice Chat
Group Discussion	What's difference?	physical appearance, kinesics-gestures, body movement	Avatar Video Slides
Virtual Field Trip	Hide & Seek	do.	Text /Voice Chat
	Exercise & Explore		AWEDU/ Second Life

Research Methods

Our ultimate project goal is to answer the main research question: How can multimodal communication in the CVE context to improve the EFL learner's communicative competence and stimulate their interaction? Investigations of the potential for applying 3D virtual reality to develop communicative competence have been continuously conducted to this day. The results provide practical implications for future application possibilities of gesture integration in English language learning. Graduate and undergraduate students participated in both real-world and virtual activities, such as navigation, on-line discussion and team work.

This study is primarily a qualitative investigation employing the ethnography of communication (Saville-Troike, 1982; Hymes, 1972, 1992; Schiffrin, 1994) to explore speech acts and communicative event systems within a virtual context. The

virtual community is regarded as a virtual community of practice in this study. Two graduate students and three undergraduate students, ages 20-24, participated in our study. Students volunteer to join the project and need to meet the prerequisite, the Intermediate Level on the General English Proficiency Test (GEPT). During one year, they participated in both physical and virtual events. The instructor and students gathered in VEC3D and spent 1.5 hours per week participating in the virtual events. Text-based chat logs, voice chat records, and video materials were collected for the follow-up communicative interaction analysis consisting of content and speech linguistic discourse analyses to unveil social meaning of speech acts, and provide a description of communication strategies and forms in multimodal CVEs.

Results and Expectations

The innovative combination of multimodal communication forms increased the EFL learners' autonomy to call CSs into action. The different types of CSs occur coincidentally or by turns. The increasingly frequent interplay between verbal and nonverbal communication strategies was found in the study. Communication strategies of non-linguistic means, fillers and gambits were used frequently under real-time circumstances. The overall interaction, communication, and usefulness of CVEs were enhanced by the multimodal system which yields more options for learners. The multimodal system makes it possible to reduce the restrictions and limitations of the traditional CVEs.

The preference for communication with VEC3D is mainly voice-based due to the time pressure and communication naturalness. When learners conduct text-based communication, participants, naturally, have the tendency of concentrating on the text chat window, and shifting their attention from the CVEs. Basically, the CVE systems, such as Second Life and Active Worlds, allow users to operate two modes- chat and avatar movement control - or more. When the text mode demands user attention, thus it distracts users from the other modes. Users usually become overloaded with visual, multimodal information and communication when operating two or more modes at the same time. It was evident from the data that voice-based communication is preferable over written communication, especially when real-time response and concentration on avatar controlling in CVEs are required.

The acquisition of nonverbal communication and paralinguistic communication strategies in the CVEs was evident in the data. The aspects concerning the nonverbal communication analysis in the current study involve avatar physical appearance observations, gestures and body movement, facial expressions and paralanguage. The participants were allowed to select the avatar's physical appearance from a menu. To

make recognition easy and reveal their identity in a small group setting, they preferred to select an avatar that has some similarity to their own appearance such as hair style and clothes. Many conversation topics relating to the avatar's physical appearances were found. The avatar's physical appearance naturally and intuitively allowing learners to address ideas about identity, personality and creativity was used to draw attention to themselves.

Strategic competence includes non-linguistic strategies used to foster communication. Savignon indicated, "A gesture may serve as a coping strategy by either filling in for a word or expression or sustaining rapport throughout a momentary silence" (1983, p. 44). Body language is included among the communication strategies used by the learners through miming. For instance, in the early stages of CVE communication, the act of waving the hand and arm as a greeting and attention-gaining device is used to replace the word, Hello. Unexpected communication problems can frequently be solved using emoticons or text smileys to convey meaning and emotions when writing and using body language while talking. The learners employ nonverbals to promote communication. For example an avatar named Kou uses a walking gesture to show how to reach the target location, 'waterfall'.

Wen: *hello*

Kou: *hello ...(She shouted) there's a waterfall over there*

Wen: *how to get there?*

Kou: *follow me (avatar walking through and forward)*

This study shows the acquisition of strategic competence, multimodal communication process and interaction in the virtual learning community. The CVEs provides the multiple communication channels for learners to overcome communication breakdowns more easily. Data collected from the interviews and participant observation illustrates the learners relied heavily on non-linguistic means, such as mime, and manipulating virtual objects. The use of 3-D Virtual Reality can play a crucial part in providing an optimal context for the foreign language learners to acquire the strategic competence that ensures that the conversation remains ongoing. Moreover, the communication gap between interlocutors is more successfully bridged via multimodal VEC3D. Authentic context, synchronous text-based and voice-based communication, language-based and nonverbal communication all play critical roles in natural language acquisition and the acquisition of strategic competence.

The research conducted by Manninen (2002) is relevant to this study as it intends to provide multiple users with various interaction forms, including language-based and nonverbals, to enhance communication, and analyze multimodal communication events in CVEs through using interaction analysis. Based on Manninen's concept model of interaction forms in CVEs, we developed a concept model of Communication Strategies (CSs) (See Figure 4.) for the future description and analysis of communication forms and CSs in the CVEs.

The concept model has been constructed to conceptualize CSs in the context of the CVE. CSs are either verbal or nonverbal. On the right side of Figure 4 are main categories (paraphrase, borrowing, appeal for assistance, avoidance and others) containing types of verbal CSs based on literature reviews, e.g., Tarone's (1981) taxonomy of CSs, and the previous findings (Shih & Yang, 2008). On the left are the main communication channels (paralanguage, body language, personal-interpersonal space, object communication) of nonlinguistic strategies based on theoretical knowledge and related works of nonverbal communication forms (e.g., Hymes, 1972; Argyle, 1975; Laurel, 1993; Manninen, 2000). The model serves as the conceptual framework for analyzing communication forms of CSs in the study.

The communication forms have been largely text-based and language-based in CVEs. The lack of nonverbal communication clues supporting social interaction and sophisticated communication is obvious in text-based chat and today's CVEs. With the advent of synchronous voice-based communication tools, communication shift from text to talk in multiplayer games (Halloran, Rogers & Fitzpatrick, 2003), and CVEs, such as Voice Chat in Second Lifeⁱ and VEC3D, is in progress. Possible benefits may include increased communication naturalness and synchronicity, visual and paralinguistic clues.

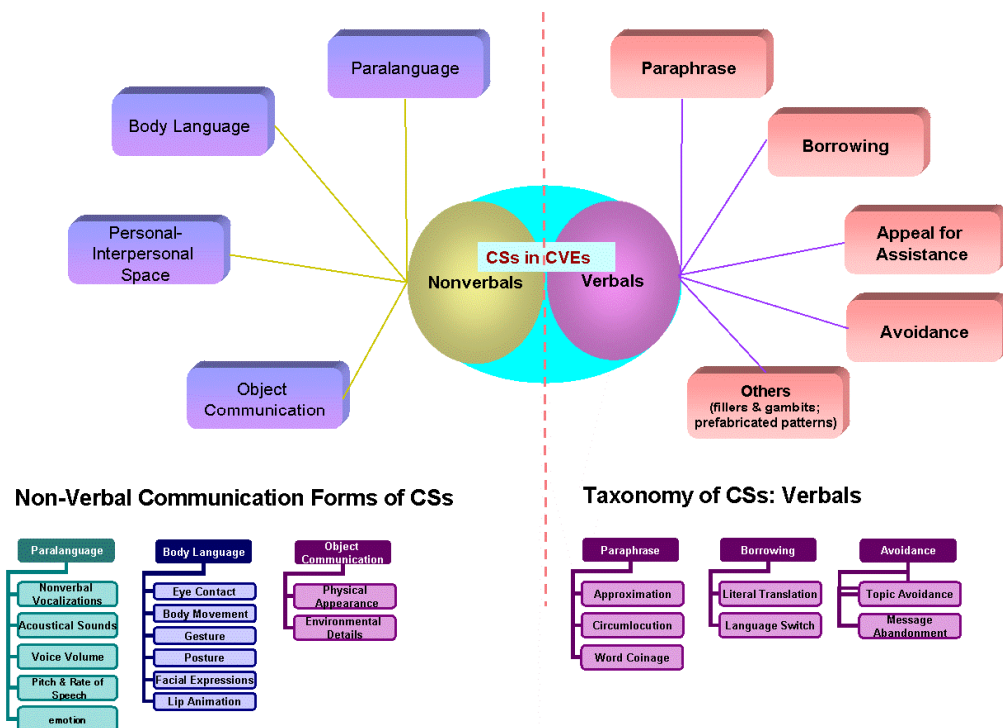


Figure 4. Concept Model of CSs in CVEs

VEC3D aims to reinforce authentic communication by offering a communicative context in CVEs. Learners are expected to overcome communication barriers and breakdowns, and develop communicative competence. Current analysis provides implication for the follow-up design and forms a tentative concept model depicting multimodal communication forms in CVEs. In addition to continuing to fulfill the object, development of communicative competence, we are beginning to elaborate our study more on the user socio-cultural experience in the forming a virtual community of practice in the multimodal CVE setting.

Acknowledgment

This research is supported by the National Science Council, Taiwan.

Copyright Statement

Copyright to this paper archived on ISTE’s Web site remains with the authors

References

Allen, L.Q. (1999). Functions of nonverbal communication in teaching and learning a foreign language. *The French Review*, 72, 469-480.

- Argyle, M. (1975). *Bodily communication*. New York: International University Press.
- Canale, M. (1983). From communicative competence to communicative language pedagogy. In J. C. Richards & R. W. Schmidt (Eds.), *Language and communication* (pp. 2-27). New York: Longman.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Chen, T., & Rao, R. (1998). Audio-visual integration in multimodal communication. *Proceedings of the IEEE*, 86, 837-852.
- Crookall, D., & Oxford, R. L. (Eds.). (1990). *Simulation, gaming, and language learning*. New York: Newbury House.
- Di Pietro, R. (1987). *Strategic interactions*. Cambridge: Cambridge UP.
- Halloran, J., Rogers, Y., & Fitzpatrick, G. (2003). From text to talk: Multiplayer games and voiceover IP. *Proceedings to Level Up: 1st International Digital Games Research Conference*, 130-142. Utrecht: University of Utrecht.
- Hymes, D. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics*. Harmondsworth: Penguin Books.
- Hymes, D. (1992). The concept of communicative competence revisited. In Putz. M. (Ed.), *Thirty Years of Linguistic Evolution*. Philadelphia: Benjamin.
- Krashen, S. (1985). *The input hypothesis*. Oxford: Pergamon Press.
- Laurel, B. (1993). *Computers as theatre*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc.
- Manninen, T. (2000). Interaction in networked virtual environments as communicative action: Social theory and multi-player games. *Proceedings of CRIWG2000 Workshop*, 154-157. IEEE Computer Society Press.
- Manninen, T. (2002). Interaction forms in multiplayer desktop virtual reality games. In S. Richir, P. Richard & B. Taravel (Eds.), *Proceedings of VRIC2002 Conference*, 223-232.
- Massaro, D. W. (1998, December). Illusions and issues in bimodal speech perception. *Proceedings of Auditory Visual Speech Perception '98*, 21-26.
- Massaro, D. W., & Stork, D. G. (1998). Speech recognition and sensory integration. *American Scientist*, 86, 236-244.
- Renninger, K. A., & Shumar, W. (Eds.). (2002). *Building virtual communities: Learning and change in cyberspace*. Cambridge: Cambridge University Press.

- Richards, J. C., & Schmidt, R. W. (1983). *Language and communication*. London: Longman.
- Savignon, S. J. (1983). *Communicative competence: Theory and classroom practice*. Reading, MA: Addison-Wesley Publishing Company.
- Saville-Troike, M. (1982). *The ethnography of communication: An introduction*. Cambridge, MA: Basil Blackwell.
- Schank, R.C. (1992). *Goal-based scenario* (Technical Report No. 36). Evanston, IL: The Institute for the Learning Sciences, Northwestern University.
- Schank, R. C. (2003). *Every curriculum tells a story*. Story-centered curriculum white paper. Retrieved June 8, 2008, from the Socratic Arts Web site: http://socraticarts.com/additional_links.html
- Schiffrin, D. (1994). *Approaches to discourse*. Cambridge: Blackwell Publishers.
- Shih, Y. C., & Yang, M. T. (2008). A collaborative virtual environment for situated language learning using VEC3D. *Educational Technology & Society (SSCI)*, *11*(1), 56-68.
- Swain, M. (1984). Large-scale communicative language testing: A case study. In S. J. Savignon & M. S. Berns (Eds.), *Initiatives in Communicative Language Teaching*. Reading, MA: Addison Wesley.
- Tarone, E. (1981). Some thoughts on the notion of communication strategy. *TESOL quarterly*, *15* (3), 285–295.
- TESOL. (1997). *ESL standards for Pre-K-12 students*. Alexandria, VA: Teachers of English to Speakers of Other Languages.

ⁱ Many residents use Skype to voice chat in Second Life and Linden lab integrated voice chat into Second Life on February 27, 2007. Information relating to voice chat in Second Life available at http://lindenlab.com/pressroom/releases/02_27_07