International Journal of Linguistics and Communication
December 2017, Vol. 5, No. 2, pp. 40-52
ISSN: 2372-479X (Print) 2372-4803 (Online)

Published by American Research Institute for Policy Development
DOI: 10.15640/ijlc.v5n2a4
URL: https://doi.org/10.15640/ijlc.v5n2a4

The Combined Use of Audio and Visual Media in Computer Conferencing towards Equality in Asymmetric Second-Language Conversation

Hiromi Hanawa[1], Xiaoyu Song[1], Tang Mengyuang[1] & Tomoo Inoue[2]

Abstract

The purpose of this study is to explore and measure the impact on combined multimedia for Non-Native Speakers (NNS) to communicate with Native Speakers (NS).This article presents the effects of the combined use of audio and text, and substantiate increased participation of reciprocal peer interaction between NS and NNS. In this study, conversations were conducted in our lab between NS and NNS. We tested our predictions about participants' speech, verbal and semantic knowledge after conversations throughout the experiment. Data found that the use of audio and text increased the participants' mean length of utterance, and knowledge sharing between NS and NNS. The results indicated the impact of the combined use of audio and text on an interactive real-time discussion and positive effects on users' memory retention. This research suggests potential development in audio or video conferencing tools which incorporate interaction between people with linguistic boundaries.

Keyword: Communication, second language, non-native speaker, text-enhanced audio conference

1. Introduction

Prior research has tackled communication problems and shows difficulty in non-native speakers working together in online meetings. Non-native language speakers frequently fail to establish trust and collaborate effectively (Tenzer, Pudelko, & Harzing, 2014). When Native speakers (NS) talk with Non-native speakers (NNS) in a common language, they encounter interactional problems due to discrepancies in fluency level (Kurhila, 2001). As such, the structure of their communication is unbalanced and asymmetric. Computer-mediated communication (CMC) research has examined this in relation to online meetings working on more balanced participation with better quality of conversations (Chun, 1994; Hampel & Hauck, 2004). Thus, conferencing tools using either audio or audiovisual generate more balanced participation between NS and NNS and has become a popular research topic.

While only few studies focused on audio conferencing tools, there is research discussing the development of video interaction. It might be explained by users' emotional preferences of video communication tools (Pye & Williams, 1997). Focusing more on content analysis, video interaction research deals with tools which incorporate video annotation, browsing, editing, navigation, recommendation and summarization rather than tools that pursue collaborative multimedia interaction (Schoeffmann, Hudelist & Huber, 2015). Audio conferencing, on the other hand, is categorized as voice-oriented or screen-oriented having components of screen sharing, shared whiteboard use, presentation tool and so forth (Ding, Erickson, Kellogg, Levy, Christensen, Sussman, Wolf, and Bennett, 2007). Literature shows that audio conferencing omits non-verbal signals that may not have critical roles in business communication due to their redundancy and less constant meaning (Pye & Williams, 1997). It is noted that in business situations global project managers do not consider visual contact during online meetings to improve trust on global projects and rarely use video conferencing (Binder, 2009). Research uptakes failed to prove that video systems are more successful in business computer-mediated communication than audio ones (Echenique, Yamashita, Kuzuoka, & Hautasaari, 2014; Pye & Williams, 1997). Hence, this paper aims at exploring media effects of audio and visual tools in conferencing system.

[1] Graduate School of Library, Information and Media Studies, University of Tsukuba, Ibaraki, Japan

[2] Faculty of Library, Information and Media Science, University of Tsukuba, Ibaraki, Japan

Conversations between NS and NNS may pose potential questions on how speakers take advantage of CMC during real-time interaction. It also tries to understand whether audiovisual tools would become more efficient in real-time discussion when NS talk with NNS. Most conferencing systems use audiovisual channels (Binder, 2009), where participants share visual images and exchange their messages with other participants. However there is no quantitative data analysis that would prove actual usefulness of communication aid and provide empirical evidence that NS and NNS equally participated in interactive discussion. It is thus unclear how the combined use of audio and text affects conversation between NS and NNS. The aim of this study is to investigate how NS and NNS improve participation in online discussion and understanding the conversation content. In this study NS were asked to manually input keywords from essential portions of speech and the NS assumed to be difficult for NNS to comprehend during the conversation to support their oral speech. We conducted a lab experiment to investigate speakers' typing to the essential parts of conversations during audio conference using second language.

2. Literature Review

Most previous research (Novinger, 2001; Tenzer, Pudelko, & Harzing, 2014) refers to people of different languages embracing communication obstacles, which deliver misunderstanding and ineffective communication in business and social situations. Language is an important factor in communication. On such occasions they are mutually incapable of performing effective communication and thus often result in conflict. Since NS and NNS create unbalanced and asymmetric structure in the second language conversation, NS sometimes corrected NNS way of speaking (Kurhila, 2001). Literatures pointed how difficult the multilingual communication is, and suggested computer-mediated communication (CMC)to balance this with more equal participation and quality of conversations(Chun, 1994; Hampel & Hauck, 2004). Developing these communication systems featuring more balanced participation, interdisciplinary research has become a popular research topic across the field of human computer interaction (HCI), multimedia, information processing and second language learning.

Some multimedia combination has improved group conversations especially in participants' ability of comprehension and contribution on conversations. Hamplel and Hauck(2004) developed an online conferencing system using audio and visual channels, and Chun (1994) reported that speakers demonstrated alternative discourse competence via computer network-based interactive discussion. Additionally Warschauer (1996) found that electronic discussion allows more equal participation by comparing face-to-face and video chat. These researche rnote that CMC could improve participants' speech in second language. Other works suggest using multiple modalities of communicationto be efficient concepts. For example, Chapanis (1976) tested several channels of communication (e.g. face-to-face, voice, handwriting, and typewriting) to solve several conversation tasks. Results showed that the problem was solved faster when using a voice channel. And also Echenique, Yamashita, Kuzuoka, and Hautasaari (2014) tested two channels of communication using video and text support on multiparty audio conference to solve a conversation task. During this experiment, task performance, gestures, and common words used by three people were measured. Results showed that task performance and common words were not significantly different between video and text communication. Thus some multimedia combination may demonstrate positive effects on real-time interactive discussion when including NS and NNS, whereas Menne and Menne (1972) pointed participants' individual differences in level of education and maturation may affect result of multimedia combination.

Unfortunately, the existing multimedia comparisons are mostly based on video, and little has been done how audio enhanced communication. Pyeand Williams (1997) explored uses of teleconference tools of video and audio, and analyzed types of multimedia use depending on the purposes and situations. Other work tested face-to-face and video conferencing system for group collaboration in which users show ability to express understanding and attitudes (Isaacs& Tang, 1994). To show a wide variety of research, Schoeffmann, Hudelist & Huber (2015) reported summary of research that incorporates video interaction tools found it works over the last few years have less focus on collaborative video interaction that need users' joint efforts. As such, neither studies have provided audio conference was improved by the use of graphic channels for users' collaboration, nor audio conferencing by different native languages have yet to be conducted. Ding, Erickson, Kellogg, Levy, Christensen, Sussman, Wolf and Bennett (2007) shows that audio conferencing has two types: voice-oriented or screen-oriented, and investigated the advent of voice over internet (i.e. VoIP) with visual channel, which improved users' calling experience. Audio conference omits users' non-verbal information, and utilizes graphic channels (e.g. shared documents, chat screen, presentation tools).

To improve sound quality of conference calls, Yamashita, Echenique, Ishida & Hautasaari (2013) examined effects of transmission lags by insertion of artificial delay during audio conferencing, and reported optimal length of lag. These studies show voice-oriented approaches managed sound quality, however few screen-oriented methods had investigated how audiovisual channels enhanced audio conference.

Although little work has investigated audio conferences, some studies show that text enhances group communication. Yankelovich, McGinn, Wessler, Kaplan, Provino & Fox (2005) developed text chat screen to investigate the effect of private communication during online group meetings. Other works also reported group collaboration has been improved by converting speech-to-text transcript which caused a variety of different effects on conversations. As for current technological support for group communications, researchers explored Automated Speech Recognition (ASR) and Machine Translation (MT). However ASR and MT create inaccuracies in transcripts that impair NNS comprehension (Yamashita, Inaba, Kuzuoka, & Ishida, 2009). According to the prior studies, human volunteers created more accurate transcripts than ASR during the real-time caption processing (Lasecki, Kushalnagar, &Bigham,2014). To improve the accuracy of ASR transcripts, ASR with a human volunteer editing transcript was tested and demonstrated positive effects on conversation among group of people engaging in a collaborative task (Gao, Yamashita, Hautasaari, Echenique, & Fussell, 2014). Another experiment tested whether NS highlighting on ASR transcript help NNS, and found it beneficial for NNS to communicate with NS in real-time conversation (Pan, Yamashita&Wang,2017). These studies showed how speech-to-text transcript demonstrated positive effects on conversations, although erroneous transcript increases burden of NNS when talking with NS. In this regard, NNS burdens may be explained as researchers stated that processing the second language increases the cognitive load. Second language listeners often face real-time listening difficulty and quickly forget what is heard, as well as unable to form the next possible sentence that intends full content of messages and ideas (Goh, 1999). Moreover, reading transcripts during talking adds cognitive loads and multitasking skills in cases the conversation content is dense, and NNS are forced to read long transcripts (Warschauer, 2000).

To reduce NNS load of reading transcripts, we called on NS to type keywords from essential portions of speech and the NS assumed to be difficult for NNS to comprehend during talking. Communication tools that speakers employed typing message, statement, or essential part in conversation may deliver useful content when used in computer teleconferencing. Human volunteer editing is common and a beneficial way for NS andNNS to cope with uneven language proficiency (Echenique, Yamashita, Kuzuoka, & Hautasaari, 2014). Our prediction is that participants use two modalities of listening and reading to perform efficient interaction. In conversation between NS and NNS participants need to adjust their speech to accommodate for communication boundaries. Text input while speaking may pose challenges for NS and at the same time reduce language barrier between NS and NNS (Inoue, Hanawa & Song, 2015). Transcript that NS types does not show a whole sentences from speech but produces keywords of conversations considering mental and physical load on NNS while typing task was imposed. Typed words would become reasonable speech accommodations that help NNS to understand NS speech. Showing essential portions of speech, human editing transcripts may become more helpful for NNS than ASR transcripts that present entire conversation transcripts.

The purpose of this paper is to show how NS and NNS managed audio conference with the combined use of audio and text in conversations. Our study is motivated by earlier research: how CMC helped balanced participation on discussion(Chun, 1994; Hampel& Hauck, 2004), and how CMC improved participants' memory of the conversation that speakers understood conversation content from speech (Najjar, 1998; Nasser& McEwen, 1976; Pye & Williams, 1997). To date, a number of literature showed computer-mediated communication (CMC) that has audio, whiteboard, and shared document presentation led to more equal students' participation in second language classroom, whereas face-to-face discussion tends to create dominant individuals who determine the conversation topic working in small groups (Hampel & Hauck, 2004). On such conversations, CMC offered students' spoken discourse competence and oral language acquisition due to audio-graphic component that provide flexibility and self-pacing (Hampel & Hauck, 2004). In other words, CMC presented a number of students' different interactional speech act in terms of sociolinguistic and interactive competence (e.g. ability to express, interpret, expanding the topic, negotiating meaning, and advocating the turn-taking) (Chun, 1994). Thus we assumed that audio and text increased speakers' participation due to alternative speech act when used in conversations between NS and NNS. We also investigated how participants exchanged information from real-time conversations to retain their mutual knowledge. Literature reported CMC presents positive effects on learning through multimedia combination and interactive user interface, as some multimedia combination helps people to learn from text, graphics, sound and video (Najjar, 1998).

Previous work examined children learning and found that combination of audio-text presentation was better than text-only and audio-only conditions by words recall test of presented videotapes (Nasser& McEwen, 1976). Other work investigated students learning performance that was significantly greater in combined stimuli of tape-recording and related photographs than stimuli of tape-recording and unrelated photographs (Severin, 1967). These studies suggest that some multimedia combinations are advantageous than others, such as single media for educational multimedia user interfaces. This concept is called "cue summation principle", which would be useful in actual communication aid tools and provide information that people understood from multiple sources of information. Thus we predicted that speakers improved memory of interaction with the combined use of audio and text. In this paper we examined participants' memory of conversations that represent comprehension of the conversation content.

3. Methods

3.1 Overview

Aim of this paper is to investigate how NS and NNS improve participation and understanding in conversation with the use of audio and text that NS type keywords from their messages. We executed the conversation experiment under two conditions by audio and audiovisual. Experimental design was within subjects. Conversation tasks and its order were balanced between subjects. 16 pairs participated in both conditions within a day. This experiment suggests the use of audio and text as communication media that present actual usefulness when supporting interactive discussion. Result shows empirical evidence of speakers' equal participation toward real-time discussion and contributions.

3.2 Participants

Participants in the experiment were 16 pairs of Japanese and Chinese. NS were Japanese participants who were born and grew up in Japan. NNS were Chinese participants who were international college students from China. NS and NNS were paired according to their availability, none of them knew each other before the experiment. All participants were volunteers. The gender distribution was 17 male and 15 female, and their mean age was 25.3. Computer literacy among participants were not greatly varied, and all had one minute practice sessions of using computers to ensure that participants feel comfortable before data collection began. NNS participants' Japanese Language Proficiency Test (JLPT)(2017) were N1 level successful or equivalent, and their average score were 118.9 which means they have ability of listening and reading comprehension in Japanese.

JLPT (2017) is the largest-scale Japanese language proficiency test that has more than 610,000 examinees around the globe for academic and employment purposes. JLPT offers a certificate in the areas of listening and reading. As the largest-scale test, the JLPT most likely to measures participants' listening and reading comprehension of Japanese. Unfortunately, there was not an accurate indicator to test NNS communicative competence in actual conversation. Survey before the experiment showed NNS participants rated their level of speaking competency as on average 3.5 out of 7 point Likert scales of Very poor=1, Poor=2, Fair=3, Good=4, Very good=5, Excellent=6, and Exceptional=7. NNS have been learning Japanese for four years on average.

3.3 Software and equipment

The pair was seated at the PC tables in the laboratory as shown in Fig 1. This environment simulated audio conferencing that continuously deployed audible space and prevented mechanical noise. Participants utilized Lenovo B590 15.6 inch laptops with an external monitor and keyboard. Participants wore microphones to record their voice. Fig 2 shows participants' PC monitor. Activating MS Word 2013 and placing on the left side of the monitor, NS typed on it. The PC was connected to an intramural local area network. Each PC launched Skype and started a voice call. Skype is a software application commonly used for voice and video call. Monitors were synchronized with the other PC via Skype. Skype screen-sharing option hardly caused delay when synchronizing two monitors (Sirintrapun, 2012). Turning Skype sounds off, participants talked directly while conversations. Typed materials were shared with the co-participant through the network. Task-oriented information was opened on the right hand side of the monitor, and participants did not share it with the co-participant.

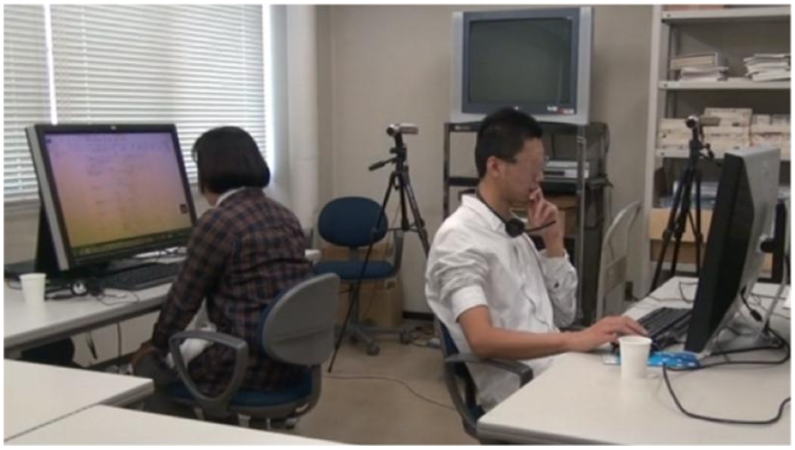
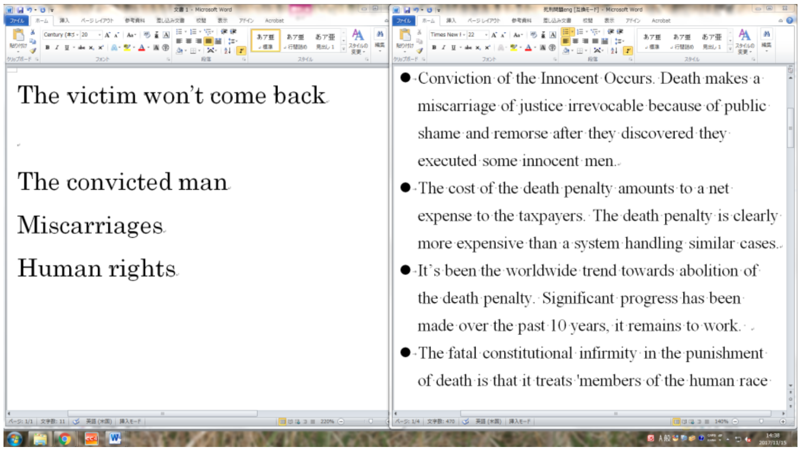


Fig 1 Experiment equipment.



The victim won't come back

The convicted man

Miscarriages

Human rights

- Conviction of the Innocent Occurs. Death makes a miscarriage of justice irrevocable because of public shame and remorse after they discovered they executed some innocent men.
- The cost of the death penalty amounts to a net expense to the taxpayers. The death penalty is clearly more expensive than a system handling similar cases.
- It's been the worldwide trend towards abolition of the death penalty. Significant progress has been made over the past 10 years, it remains to work.
- The fatal constitutional infirmity in the punishment of death is that it treats 'members of the human race

Fig 2 Participant's PC monitor.

3.4 Task and experiment design

The conversation tasks were a debate of the nuclear energy and death penalty system. Conversation tasks should be goal-oriented, and designed in order to convey online meeting where participants could work together and resolve real-world issues on the use of multiple sources of information. Pellettieri (2000) stated that task difficulty is an important in the design as those tasks involve vocabulary beyond the participants' repertoire, ideas, and concepts that can facilitate linguistic activity to convey messages and negotiation of meaning. From this objective, debating was beneficial for participants' interaction as well as obtaining information from one another. Generally requiring time to prepare before debating, participants took one minute to read the task-oriented information that informed major issues to educate themselves before data collection. Fig 2 shows task-oriented information. Participants modified the provided information on PC according to their own opinion, although editing was not mandatory. Participants did not see the task information composed by the co-participant. There was no moderator of debate. The conversations of the two different conditions were compared. Fig 3 shows the flow of information processing in both conditions. Participants retrieved information only from listening in the Control condition, and they utilize listening and reading input at the same time in the Keyword condition.

- •NS typed essential portions of speech and that NS assumed to be hard for NNS to comprehend on a computer keyboard while NS was speaking (Keyword condition).
- •Nobody typed on a computer keyboard and NS and NNS performed audio communication (Control condition).

According to Hirai (2003), information retrieved from listening is directly processed for comprehension, and information from visual input has a different route to reach understanding. Keywords show essential portions from speech and that NS assumed to be difficult for NNS to comprehend in real-time conversations. Hence typed texts may differ amongst all individuals.

A within-subject was adapted. Each pair participated in two conditions within a day. The combination of conversation tasks and experiment conditions were balanced to cancel out an order effect. Two conversation tasks were combined alternately with two conversation methods across pairs. Eight pairs conducted the Keyword condition first, and the rest eight pairs participated in the Control condition first.

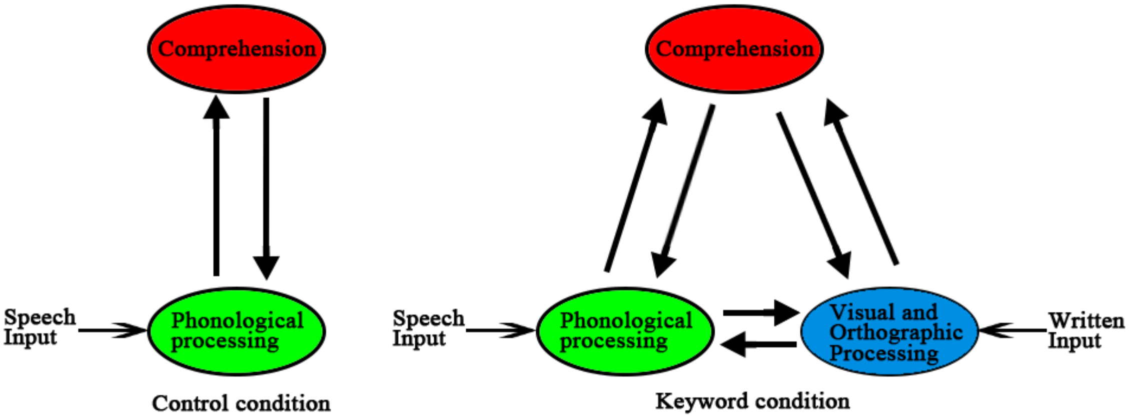


Fig 3 Phonological and orthographic word recognition process (based on Hirai 2003).

3.5 Procedure

The experiment was conducted in a laboratory, where participants held two seven-minute conversations in pairs. 16 pairs of NS and NNS participated in this experiment. Fig 4 shows experiment procedure. Participants were seated and filled out a consent form and a demographic survey which asked age, nationality, gender among other questions. Researchers provided participants with written consent forms prior to the experiment. Participants read and agreed with it, and took part in a lab experiment. All experiments were conducted in compliance with protocols reviewed by the ethics committee and approved by the University of Tsukuba. Participants chose the pros. and cons of the topic in pairs.

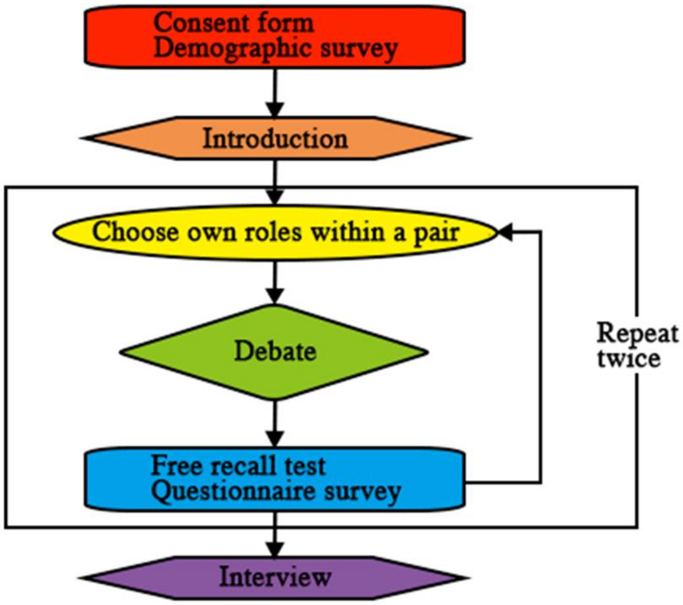


Fig 4 Experiment procedure.

4. Measures

We have two measures of analysis to explore our predictions about speakers' participation on discussion and comprehension of conversations from quantitative data. The experiment was videotaped to observe how speakers communicated with uses of audio and text. Experimenters retrieved quantitative data of participants' speech length.

Free recall tests are a measure of memory where participants study a list of items on each trial, and recall the items in any order (Tanaka, 1998). In the experiment participants contribute to a debate and recall and present words and phrases in any order after each round. As for qualitative data, interview assessments were obtained to offer credible proof of participants' perception by means of semi-structured interview. We positioned a video camera to record the entire laboratory room, and two video cameras from the participants' right hand side in order to capture their upper torso including typing behavior. Participants' PC monitor was recorded with the desktop capture software. Obtained data were video, screen capture, typed characters on the PC, results of free recall tests, questionnaires, and interview comments.

4.1 Speech length in conversation

Speech length was measured in order to investigate how NS and NNS increased participation in discussion with the use of audio and texts. Collected video data through the experiment was 224 minutes as a whole. Video data was combined in EUDICO Linguistic Annotator (ELAN), which is a text annotation application to create an embedded text annotation for the media files. Annotated texts in every single timelines are saved with the time corresponding to the annotation (Rosenfelder, 2011). To obtain each pair speech length, speaker's utterance was described using ELAN with inter pausal unit (IPU). IPU is a pause-free unit of speech from a single speaker separated by a pause at least 50 ms, marked out with pose (Levitan, 2011). Results show objective evidence based on quantitative data analysis about participants' speech length. Each conversation was 7 minutes for both conditions, thus experimenters measured speech length uttered in 7 minutes but not included utterance after 7 minutes passed for this measure.

4.2 Free recall test

Free recall tests are a test of memory that participants brought back after conversations in any order (Tanaka, 1998). We conducted the free recall test to investigate whether NS and NNS understood the conversation as well as successfully shared and retained the same information through the conversation with the use of audio and text. Participants recorded information from the conversation using only their memory. Participants individually typed on PC after the conversation by words, phrases, or sentences as much as they could. We divided phrases and ideas with spaces and the line breaks participants inserted, and counted the number of idea units under each condition. Thereafter we counted the number of matching words and phrases that occurred in the summaries by both participants. Table 1 shows criteria and examples of matched words and phrases. Words and phrases with similar connotations were counted as a match. For example, "Prison" is "a building in which people are held as a punishment", "a place of confinement especially for lawbreakers", or "a state of confinement or captivity" hence these are identical as people generally recognize the same in meaning. "Antarctica" and "baseball" are not identical in meaning as seen the difference in meaning (Tanaka, 1998).

Table 1 Criteria and examples of matched words.

Category	NS	NNS
The same words	Great east Japan earthquake	Great east Japan earthquake
	Economic efficiency	Economic efficiency
	False charge	False charge
	Radioactivity	Radioactivity
Identical in meaning	Nuclear accident	Accident at nuclear power plant
	Risks of nuclear power generation	Nuclear power generation is dangerous
	Alternative sentences of death penalty	Other way to bring criminals to justice
	Recidivism risks	Repetition of the offense after prison
Mismatches	Good fuel efficiency compared to thermal power	Saving natural resources
	Risk of false accusation	The death penalty cost less than life in prison
	Few people are technical experts	People's loss of profit
	Risks of transport of radioactive materials	Not good for marine environment

4.3 Semi-structured interviews

Semi-structured interviews provide qualitative data of participants' feedback about experiment and uses of multimedia combination. The interview questions were developed through pilot interviews with senior researchers, and covered five areas such as own communication, peer's communication, mood, frustration and technology.

Researchers conducted interviews in respondents' dominant language either of Japanese or Mandarin Chinese. Predetermined questions devoted to inquiring about the use of multimedia and impression of work context.

5. Results

5.1 Speakers' utterance

Audio and text presentation was expected to facilitate participants' speech when NS provided text transcripts to cope with different fluency level to NNS. Researchers measured speakers' speech length in order to investigate how NS and NNS increased participation on discussion. Fig 4 indicates mean length of utterance that is average length of every single utterance of NS and NNS. We ran a two way repeated measures ANOVA with the Condition (Keyword, Control) and language proficiency (NS, NNS) as within subjects factors revealed main effects of text, F(1, 4102)=119, p<0.01, and speakers' language proficiency F(1, 4102)=8.97, p=0.003. These main effects were qualified by an interaction between text transcript and language proficiency, F(1, 4102)=15.5, p<0.01. A Bonferonni post hoc test indicated that mean length of utterance differ significantly between the Keyword and the Control condition (p < 0.01), and NS and NNS in the Keyword condition (p < 0.01).

Fig6 represents the total speech length within a task conversation. Experimenters measured the total speech length of NS and NNS, which was sum of their every single speech uttered in 7 minutes of debating. Utterance after 7 minute passed was not included in calculation of the total speech length. Then we ran a two way repeated measures ANOVA with the Condition (Keyword, Control) and language proficiency (NNS, NS) as within subjects factors revealed main effects of text, F(1, 60) = 6.92, p = 0.011.

Quantitative result according to ELAN provides data that NS and NNS extended mean length of utterance and total speech length with uses of audio and text. Fig 5 indicates extended mean length of utterance for both of NS and NNS in the Keyword condition. Fig 6 shows that NS and NNS increased total speech length during a debate. Speech length was obtained from ELAN annotation with the corresponding time. Quantitative data found that speakers increased participation in conversations and reduced awkward silence in communication across linguistic boundaries by CMC.

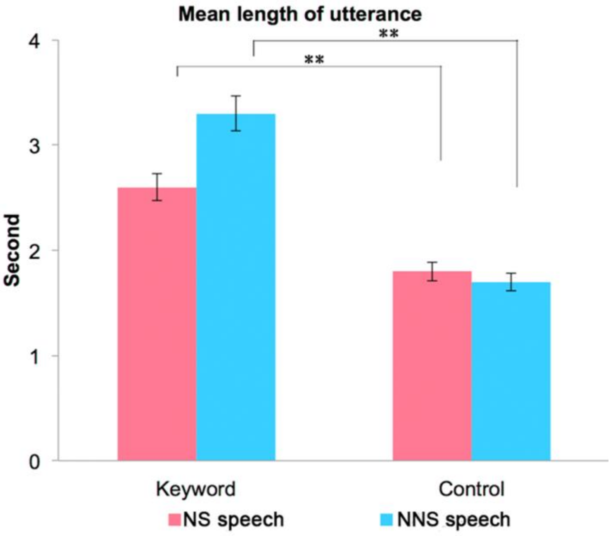


Fig 5 Mean length of utterance. N=16, **: p<.01

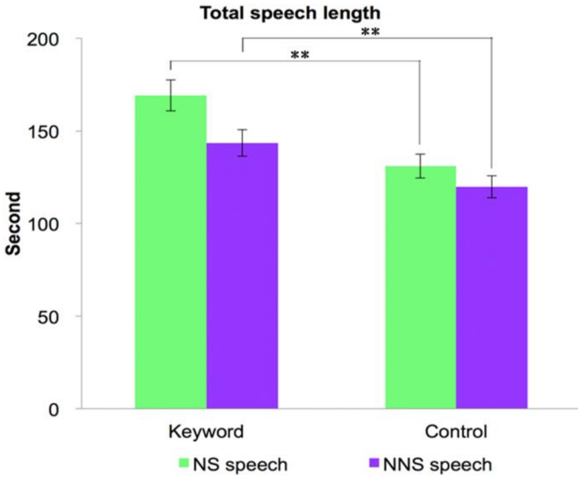


Fig 6 Total speech length. N=16, **: p<.01

5.2 Recall test performance

In order to investigate whether NS and NNS understood conversations and successfully shared knowledge on CMC, researchers conducted free recall tests that participants recorded shared information from the conversation using only their memory. The uses of audio and text were expected to enhance participants' shared knowledge in conversations due to the positive effects on learning. Fig 7 indicates the mean number of matched words that is the respondents' score averaged by pairs. Two experimenters scored independently all matched words retained by two people in common after conversations, and inter-rater agreement was high (k=0.75). We ran paired t-test on the free recall test performance. There were significantly larger matched words on the Keyword condition (M=3.8, SD=1.1) than the Control condition(M=1.8, SD=0.7), t(15)=13.5, p<0.01.

Fig8represents the mean number of idea units that participants recorded after the conversation. It is neither significantly different on NS between the Keyword condition (M=4.6, SD=1.3) and the Control condition (M=4.7, SD=1.7), nor on NNS between the Keyword condition (M=4.6, SD=1.7) and the Control condition(M=4.3, SD=1.6).

The number of words appeared in conversations were 3.8 (mean/pair) about Nuclear energy and 3.9 (mean/pair) about Death penalty system respectively. Matched words on the Keyword condition included 74% of typed information as seen on 2.4 words (mean/pair) were typed during conversations out of all 3.3 matched words (mean/pair). Matched words on the Control condition were 1.8 (mean/pair), which did not include typed words because nobody recorded them in the Control condition. Result indicated the conversation using audio and text presented not only accurate information transaction but also kept them in users' mind after conversations. Respondents largely employed typed information in free recall test depending on their memory.

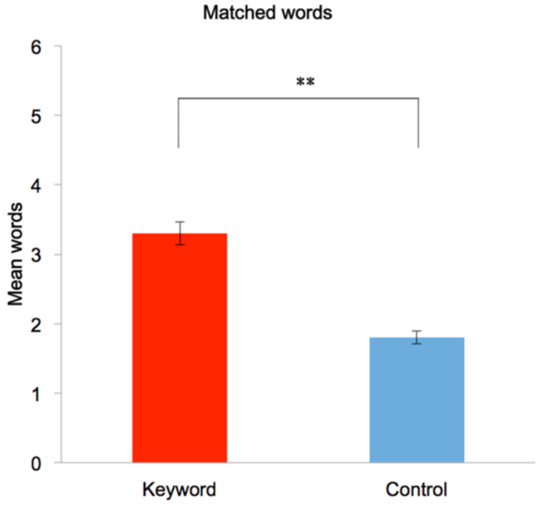


Fig 7 Mean matched words between NS and NNS. N=16, **: p<.01

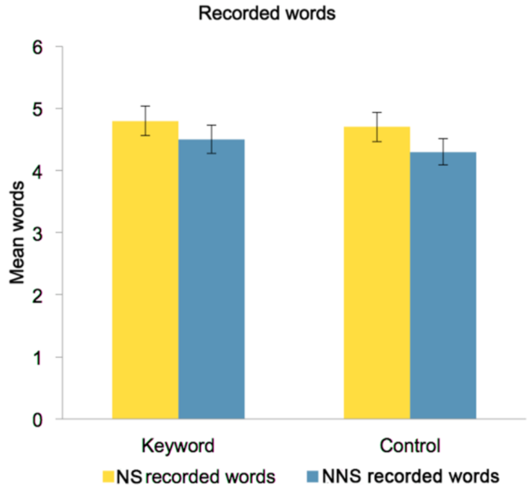


Fig 8 Mean words participants recorded.

5.3 Interviews

Semi-structured interview addressed participants' subjective evaluation from the perspective of quality of group communication with the use of multimedia. Of our 32 informants, 13 NNS and 15 NS provided positive description about the peer's communication. These comments supported the fact that audiovisual presentation improved participants' perception of communication. Most of NNS stated keywords were understandable and helpful. Furthermore, NNS testified their frustration as "texts eased anxiety that was caused by much of terminologies during debating" and "texts helped to understand what NS said." On the contrary, 12 NS remarked difficulty of own communication as "it was hard to type during talking" or "it was hard to concentrate on the conversation." These statements show that typing impose physical burden on NS, and that asymmetric load only on NS may not seem difficult for NS to take part in real-time conversations.

6. Discussion

Previous studies reported CMC in second language led students' more equal participation and provided spoken discourse competence, whereas face-to-face discussion tends to create dominant individuals in teams (Hampel & Hauck, 2004). This research investigated how NS and NNS increased participation in conversations with the use of audio and text. Fig 5 shows the mean length of speaker's utterance was significantly extended with the use of audio and text. Fig 6 also shows speakers' total speech length in 7 minutes was significantly increased by audio and text, which means speakers reduced awkward silence in conversations. Results supported "cue summation" principle that two related information is better than a single information resource. In this study NNS obtained related information from audio and texts, and the combined information was significantly superior to only audio.

Interesting finding is NNS extended their utterance more than the extent NS did. Main effects on mean speech length were qualified by an interaction between text and language proficiency. This data indicated that participants might not extend speech length by typing, but NNS increased speech by audio and text that was shared between NS and NNS. CMC presented second-language learners' different interactional speech act and interactive competence to take more active role in discourse management (Warschauer, 1996). In this study NS typed keyboard and NNS looked them. Such behavioral difference affected their speech, and speakers alteredtheir speech due to CMC. An article defined conversation tasks as linguistic activity, where NNS constructs linguistic knowledge through language use with NS (Swain, 2000). NS and NNS performed cognitive activity of verbalization to produce different speech in the target language, as well as uttered successfully to convey conversation tasks. For more details about speech act and negotiation of meaning, conversation analysis is required to show how speakers extended their speech.

The free recall tests investigated whether two people successfully shared that information and retained it after conversations. Mean of matched words between NS and NNS with the use of audio and text was larger than the only audio, as shown in Fig 7.

Other study reported the number of common references between NS and NNS was not significantly different in comparison of video and text communication in a collaborative task (Echenique, Yamashita, Kuzuoka, & Hautasaari, 2014). Our study conducted free recall test to investigate the number of matched words brought back by NS and NNS after conversations, and the number of words that NS and NNS recalled respectively. There was not significant deference in the number of words that NS and NNS remembered, but significant difference in the number of matched words associated with the conversation between NS and NNS. Two conversational tasks had the same number of keywords, and 74% of matched words included the typed information in conversations with the use of audio and text. Therefore, keywords appeared to be significant information that increased matched information between NS and NNS.

The number of matched words possibly increased because NNS not only looked at texts during the conversation but also took advantage of short pauses and vocalized them to memorize those words. Research investigated how to grasp the information in meaning chunks, as vocalization and rehearsal is effective for NNS learners' vocabulary acquisition when they utilized short pauses between passages (Hirai, 2003). Our study did not verify the effects of pauses and vocalization, so further investigation is possible to conduct regarding this point. Another reason that participants had few matched words with only audio would be word recognition that was unsuccessful in voice and speech, and those words were not retained in their memory. According to the previous work, NNS are able to identify words from their graphic form in reading, but unable to search for the meaning of spoken word when their phonological knowledge and mental lexicon differ from the acoustic input in listening due to word recognition that depends on their phonological decoding ability, lexical knowledge, and meaning retrieval ability (Hirai, 2003). NS and NNS recorded different words after speech, but decreased the word inconsistency and mismatches as well as inaccurate representations with the use of text. Parkin, Wood& Aldrich (1988) explains the relationship of listening and reading as reading text motivates learners to actively process the information more than simply hearing or repetition of the information where processing task have learners to integrate the information. When experimenters asked NS when and why did you type keywords, NS testified that they typed when noticed NNS did not understand, NNS reacted after NS typed and kept typing until NNS understood, and tried to reorganize the flow of debate. Thus speakers were more motivated to acquire information from audio and text than only audio.

7. Design implication

This study investigates how speakers with discrepancies in fluency levels demonstrate positive effects on conversation with the combined uses of audio and text. Results contribute to design implications for computer conferencing system that utilize audio and visual channel for people with different languages to engage in real-time communication. Skype Translator (Lewis, 2015) has been developed for the purpose of distance calls, where speech-to-text presentation was automated by ASR and MT. As previously shown, two communication channels were utilized and text transcript helps communication during audio conferencing and Skype chat. Our study contributed to the use of audiovisual software application, where audio provides speech accommodations to balance speakers' participation, and improve mutual understanding. Considering audio conferencing, CMC need to focus more on autographic information to improve quality of conversation instead of adding existing ASR or MT as communication tools. The combined use of audio and text is simple and feasible to conduct anywhere to support multilingual communication.

CMC could perform discourse management as well as learning through interaction in every natural language occasions across the world. Our study also contributes to suggest how CMC establish human interaction of conversation that meditates problem-based learning. Computer assisted language learning (CALL) and network based language teaching (NBLT) proposed problem-based linguistic activity contains empirical components (Warschauer, 2000), which is learning from human tutors that are a truly communicative and collaborative activity. Our findings show how to manage conference calls and interactive learning where people use different native languages. Linguistic activity meditates the learning of conceptual content such as math, science, and others, hence participants engage in knowledge building with own linguistic competence and social interaction (Swain, 2000). Our study contributes to develop tutoring systems through linguistic activities that are problem-based learning attributed to principles from a pedagogical perspective.

8. Limitation and future direction

The results of this study only show effect of the multimedia use when used in debate task problems. It does not certify the same result in all conversation tasks. These conversation tasks may have different characteristics and measures, hence the outcome of multimedia usage may change accordingly.

Moreover, when participants performed their predetermined role in conversational tasks, the conversational contents as well as their behavior would differ depending on their task roles. In this respect, the conversation tasks affect task performance and results.

As for experiment settings, participants seated back to back as a simulation environment of audio conferencing. Non-verbal information was not included. Analysis showed experimental results, but did not stretch into further description of when and what NS typed. In this study NS typed both of NS and NNS utterance, and what NS typed differs amongst all individuals. These features also may affect result of the study. We expect to explore participants' typing behavior and motivation of typing in successive research.

9. Conclusion

This research investigated the impact of the combined use of audio and text in conversation between NS and NNS. Data found that it improved speakers' participation in discussion and knowledge sharing about a task while on audio conferencing. Experiment results indicated the positive effects of audio and text as a communication medium rather than only audio. Participants highly evaluated the combined uses of audio and text as a useful way of communication, whereas NS typed essential portions of conversation and that NS assumed to be difficult for NNS to comprehend from speech. These findings provide implication towards better interaction as well as learning through communication across boundaries.

Acknowledgement

We thank our colleagues for assistance with experiment, and Dr. Kansara. H and Macapagal, S. for editorial suggestions that improved the manuscript.

References

Chapanis A. (1976), Human factors in teleconferencing systems. Final report. John Hopkins Univ. Baltimore, MD. Dept. of Psychology. National Science Foundation, Washington, DC; Available from http://files.eric.ed.gov/fulltext/ED163902.pdf

Chun, D. M. (1994). Using computer networking to facilitate the acquisition of interactive competence. System, 22(1), 17-31. doi:10.1016/0346-251X(94)90037-X

Ding, X., Erickson, T., Kellogg, W.A., Levy, S., Christensen, J., Sussman, J., Wolf, T.V. and Bennett, W.E. (2007), An Empirical Study of the Use of Visually Enhanced VoIP Audio Conferencing: The Case of IEAC, Proc. CHI '07, pp.1019–1028 doi: 10.1145/1240624.1240780

Echenique A, Yamashita N, Kuzuoka H, Hautasaari A. (2014),Effects of video and text support on grounding in multilingual multiparty audio conferencing. Proceedings of the 5th ACM international conference on collaboration across boundaries: culture, distance and technology; Japan; p. 73-81. doi:10.1145/2631488.2631497

Gao G, Yamashita N, Hautasaari A, Echenique A, Fussell S. R. (2014), Effects of public vs. private automated transcripts on multiparty communication between native and non-native english speakers. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; Toronto, Ontario, Canada; ACM; 2014. p.843-852. doi: 10.1145/2556288.2557303

Hampel, R., & Hauck, M. (2004). Towards an effective use of audio conferencing in distance language courses. Language Learning & Technology: A Refereed Journal for Second and Foreign Language Educators, 8(1), 66-82. Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.125.14&rep=rep1&type=pdf

Hirai A. (2003),The role of passage in listening instruction, The bulletin of the Kanto-koshin-etsu English Language Education Society. 17, p.113-122. doi: 10.20806/katejo.17.0_113

Inoue T, Hanawa H, Song X. (2015) With a Little Help from My Native Friends: A Method to Boost Non-native's Language Use in Collaborative Work. Proceedings of the Ninth International Workshop on Informatics; September 6-9. Amsterdam, Netherlands;p.223-226.Available from http://www.infsoc.org/conference/iwin2015/download/IWIN2015-Proceedings.pdf

Isaacs, E. A., & Tang, J. C. (1994). What video can and cannot do for collaboration: a case study. Multimedia systems, 2(2), 63-73. Doi: 10.1007/BF01274181

Japanese Language Proficiency Test (2017), Japan, [Online], Retrieved on from http://www.jlpt.jp/e/about/levelsummary.html (November 27, 2017)

Kurhila S. (2001)Correction in talk between native and non-native speaker. J Pragmatics. 33.7. p.1083-1110. doi; 10.1016/S0378-2166(00)00048-5

Lasecki WS, Kushalnagar R, Bigham JP. (2014), Helping students keep up with real-time captions by pausing and highlighting.Proceedings of the 11th Web for All Conference; April 07-07. Seoul, Korea; ACM; p. 39. 2014 doi: 10.1145/2596695.2596701

Levitan, R., Hirschberg, J. (2011) Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions. In: Interspeech, Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.416.4406&rep=rep1&type=pdf

Lewis WD. (2015)Skype translator: Breaking down language and hearing barriers, Proceedings of Translating and the Computer; Nov 26-27. One Birdcage Walk, London, UK. AsLing. p. 58-65. Available from https://www.microsoft.com/en-us/research/wp-content/uploads/2016/04/TC37-Paper-FINAL.pdf

Menne, J. M., &Menne, J. W. (1972). The relative efficiency of bimodal presentation as an aid to learning. Educational Technology Research and Development, 20(2), 170-180. Doi: 10.1007/BF02768415

Najjar, L. J. (1998). Principles of educational multimedia user interface design. Human Factors: The Journal of the Human Factors and Ergonomics Society, 40(2), 311-323. doi:10.1518/001872098779480505

Nasser, D., & McEwen, W. (1976). The Impact of Alternative Media Channels: Recall and Involvement with Messages. AV Communication Review, 24(3), 263-272. Doi: 10.1007/BF02768651

Novinger T.(2001). Intercultural Communication: A Practical Guide. University of Texas Press; Available from https://febrianafebri2.files.wordpress.com/2014/04/tracy_novinger_intercultural_communication__a_pbookza-org.pdf

Pan, M, Yamashita N, Wang. H(2017).Task Rebalancing: Improving Multilingual Communication with Native Speakers-Generated Highlights on Automated Transcripts. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. ACM, February 25 - March 01. Portland, Oregon, USA; ACM. p. 310-321. doi: 10.1145/2998181.2998304

Parkin, A. J., Wood, A., & Aldrich, F. K. (1988). Repetition and active listening: The effects of spacing self-assessment questions. British Journal Of Psychology, 79(1), 77. Doi: 10.1111/j.2044-8295.1988.tb02274.x

Pellettieri, J. (2000). Negotiation in cyberspace: The role of chatting in the development of grammatical competence. Network-based language teaching: Concepts and practice, 59, 86.DOI 10.1007/978-0-387-30424-3_105

Pye, R., & Williams, E. (1977). Teleconferencing: Is video valuable or is audio adequate? Telecommunications Policy, 1(3), 230-241. doi:10.1016/0308-5961(77)90027-1

Rosenfelder, I. (2011).A short introduction to transcribing with elan," Technical report, Univ. of Pennsylvania, January

Severin, W. (1967). The effectiveness of relevant pictures in multiple-channel communications. Educational Technology Research and Development, 15(4), 386-401. Doi: 10.1007/BF02768651

Schoeffmann, K., Hudelist, M. A., & Huber, J. (2015). Video interaction tools: A survey of recent work. ACM Computing Surveys (CSUR), 48(1), 1-34. doi:10.1145/2808796

Sirintrapun J, Cimic A. (2012) Dynamic nonrobotictelemicroscopy via skype: A cost effective solution to teleconsultation, J of pathology informatics. Aug 25. 3(1), p.28. doi: 10.4103/2153-3539.100150

Swain, M. (2000),The output hypothesis and beyond: mediating acquisitionthrough collaborative dialogue, Sociocultural theory and secondlanguage learning, pp.97-114 DOI: 10.12691/education-1-5-3

Tanaka S, Fukaya M. (1998).Development of cognitive semantics theory, Books Kinokuniya, Tokyo, Japan. p.28-35.

Tenzer, H., Pudelko, M., &Harzing, A. W. (2014). The impact of language barriers on trust formation in multinational teams. Journal of International Business Studies, 45(5), 508-535. Doi: 10.1057/jibs.2013.64 Available from https://harzing.com/download/mnt.pdf

Warschauer, M. (1996). Comparing face-to-face and electronic discussion in the second language classroom. CALICO Journal, 13(2-3), 7-26. Available from http://education.uci.edu/uploads/7/2/7/6/72769947/comparing_face-to-face_and_electronic_discussion.pdf

Warschauer, M, and Richard Kern, eds. (2000), Network-based language teaching: Concepts and practice. Cambridge university press Available from http://docshare01.docshare.tips/files/28513/285138413.pdf

Goh, C. C. (2000). A cognitive perspective on language learners' listening comprehension problems. System, 28(1), 55-75.Doi: 10.1016/S0346-251X(99)00060-3

Yamashita, N., Inaba, R., Kuzuoka, H., & Ishida, T. (2009). Difficulties in establishing common ground in multiparty groups using machine translation. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems pp. 679-688. ACM.doi: 10.1145/1518701.1518807

Yamashita, N., Echenique, A., Ishida, T., &Hautasaari, A. (2013). Lost in transmittance: how transmission lag enhances and deteriorates multilingual collaboration. In Proceedings of the 2013 conference on Computer supported cooperative work pp. 923-934. ACM. DOI: https://doi.org/10.1145/2441776.2441881

Yankelovich, N., McGinn, J., Wessler, M., Kaplan, J., Provino, J., & Fox, H. (2005). Private communications in public meetings. In CHI'05 extended abstracts on Human factors in computing systems pp. 1873-1876. ACM. DOI=http://dx.doi.org/10.1145/1056808.1057044

Supporting information

Experiment Video. Available fromhttps://www.youtube.com/watch?v=czWZURfjHDY