THE INFLUENCE OF SOCIAL CUES AND COGNITIVE PROCESSES IN COMPUTER MEDIATED SECOND LANGUAGE LEARNING

by

Janel Rachel Goodman Murakami

A Dissertation Submitted to the Faculty of the

GRADUATE INTERDISCIPLINARY PROGRAM
IN SECOND LANGUAGE ACQUISITION AND TEACHING

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2017

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Janel Rachel Goodman Murakami, titled The Influence of Social Cues and Cognitive Processes in Computer Mediated Second Language Learning and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

_____ Date: 8/3/2017
Janet Nicol

_____ Date: 8/3/2017
Linda Waugh

_____ Date: 8/3/2017
LouAnn Gerken

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 8/3/2017
Dissertation Director: Janet Nicol

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of the requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that an accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Janel Rachel Goodman Murakami

Acknowledgements

I would like to offer my sincere appreciation for the support and encouragement provided by my dissertation committee, and to all those who helped make completion of this project possible. My committee chair, Dr. Janet Nicol, has not only been supportive in her guidance throughout the dissertation process, but has also been a source of kindness and understanding during the many events that occurred during my time in the Second Language Acquisition and Teaching program.

I have been exceedingly fortunate in knowing Dr. Linda Waugh throughout both my Masters and Ph.D. programs, and am very grateful for her support. I've enjoyed the courses I've taken with her, as well as having her guidance and feedback on my research. My time in SLAT would not have happened if not for her.

I would like to also thank Dr. LouAnn Gerken, who has offered valuable feedback and support for my dissertation work. Her insight as one who works in L1 research has been very informative and has added perspective to the research that went into this dissertation.

I have deep gratitude for my husband, Yoshihiro, who not only offered loving support, but also worked as the native speaking Japanese consultant and research assistant on this project. He sacrificed many weekends to administer lessons to my participants; his patience and support are amazing.

I also wish to thank my parents, Debi and David, and sister, Karlie, whose support and understanding are the ultimate reasons for anything and everything I do, including pursuing my doctoral degree. My friends also deserve thanks, for encouraging me when things were difficult.

And finally, I would to thank and remember my dear friend and fellow graduate student, Tamara Boyens, who tragically left us too soon, before she could complete her dissertation. Part of my motivation to complete this project comes from the desire to honor her, as we had planned to finish our dissertations together. I am grateful for the support, humor, and light she provided while I knew her, and will never forget her.

Dedication

To my beloved family, Debi, David, and Karlie

And my beloved husband, Yoshihiro

Table of Contents

List of Figures

List of Tables

Abstract

This dissertation investigated the effects of technological mediation on second language (L2) learning, focusing, as a case study, on gains in listening perception of the subtle but important feature of pitch placement in Japanese. Pitch accent can be difficult to perceive for non-native speakers whose first language (L1) does not rely on pitch or tone as a distinctive feature, such as English (Wayland & Li, 2008). Pedagogically, Face-To-Face (FTF) interactions with native or near-native speakers are typically the most effective way to learn L2 sound system features due to social presence, but these interactions are not always possible because of physical distance. Mediation can facilitate these interactions, but it is unclear which type results in more learning gains. The current study compared three mediation types that vary in the information provided to the learner: audio-only (asynchronous), video (audiovisual asynchronous), and videoconferencing (audiovisual synchronous), as well as a fourth condition of videoconferencing which facilitated mutual eye contact. The lack of mutual eye contact in standard videoconferencing (due to the webcam being above the image of an interlocutor's face) can inhibit the perceived social presence (Bondareva, Meesters, & Bouwhuis, 2006). A pretest/posttest/delayed posttest design was used, which measured error rates and reaction times for a same/different discrimination task and a picture recognition task. The participants were English L1 speakers, with no prior study of Japanese. After the pretest, they received training in the form of two short lessons in beginner Japanese vocabulary and sentence building administered by a native speaking tutor, which did not explicitly address pitch placement, but used minimal pairs for this feature as vocabulary items. The lessons were followed by a posttest, and a delayed

posttest one week later. The results showed that all four conditions succeeded in improving Japanese pitch placement detection, both immediately after and up to a week after the lessons. While an ANOVA revealed no main effect of mediation type, planned comparison results suggest videoconferencing without eye contact may lead to more gains in pitch placement perception than video. A surprising suggestion by the data was that videoconferencing with eye contact may lead to worse performance than the other mediation types. An exit survey detected the self-determination of the participants, and higher self-determination correlated with worse testing performance within the videoconferencing with eye contact condition. This suggests that the addition of eye contact increased the social presence of that condition to the point that it triggered Foreign Language Speaking Anxiety (FLSA) in the participants. Overall, this study highlights that lessons and tasks administered through mediation can be used to provide native speaker input for features that are important for listening and speaking, and this can effectively help learners attend to and learn these features.

Chapter 1: Introduction

1.1 Introduction

This research uses cognitive behavior methods to investigate the effects of technological mediation on second language (L2) learning, focusing, as a case study, on gains in listening perception of the subtle but important feature of pitch placement in Japanese.

Japanese is a pitch accent language, and features minimal pairs whose only difference is that the higher pitch falls on a different mora (or CV syllable) within the words. For example, the spoken words [ha*ʃi] "chopsticks" and [haʃi*] "bridge" only differ in that a higher pitch appears on either the first or second mora (indicated here with a "*"). Pitch accent is a subtle feature, and so can be difficult to perceive for non-native speakers whose first language (L1) does not rely on pitch or tone as a distinctive feature, such as English (Wayland & Li, 2008).

Sugiyama (2006), in her study of pitch placement perception by native speakers of Japanese, noted that this feature manifests in various ways. At the lexical level, pitch accent may appear both in the rise of the fundamental frequency (F0) from its minimum to its maximum at the point of the higher pitch (Sugito, 1998), and in the difference between the absolute values of F0 maximums between two syllables (Vance, 1995). Pierrehumbert and Beckman (1988) have described this feature at the level of prosody, showing that pitch-accented syllables had higher peaks than other syllables which had phrasal peaks. Sugiyama tested controlled disyllable minimal pairs and found that pitch differences in F0 are more robust when produced sentence-medially before grammatical markers known as particles (e.g. [wa], a topic marker), than when produced in isolation.

These findings are consistent with prior research on Japanese pitch accent (e.g. Pierrehumbert & Beckman 1988; Poser 1984; Sugito 1998; Vance 1995), further confirming the idea that the feature is distinctive within native utterances. For an L2 learner of Japanese, more accurate detection of this feature could lead to more fluent interactions with native speakers both linguistically and socio-culturally.

More generally, the need to be capable of meaningful social interaction in another language is one of the primary driving motivations for many individuals who engage in learning an L2. Because of this, there has been a growing focus on defining and refining the most effective forms of interaction for the L2 learning process and environment, looking at student interactions with the instructor, fellow students, and possible speech communities. Inquiries into what types or modes of social interaction can enhance learning are growing in scope, considering both traditional and alternative methods that could possibly afford students access to social interaction in their L2. Many technological innovations, while not originally designed for pedagogical use, are being considered as potential new mediums for L2 learning in both supportive and primary roles (Ducate & Arnold, 2011). These include technologies such as text chat, video games, audio conferencing, and videoconferencing. These technologies afford "mediated" learning environments, which allow participants who are not physically co-present (or "face-to-face"; FTF) to communicate remotely through text-based, audio-only, or audiovisual integrated means. The more mediated an environment is, the more artificial or unlike a FTF encounter it is, as the interaction is more abstracted through the technology being used. Mediation can be viewed as a continuum, from highly mediated (asynchronous, in which one interlocutor is unable to see or hear the other interlocutor, and possibly know

nothing about them) to minimally mediated (synchronous, with the ability to hear and see the interlocutor's real voice and face and the interlocutors may be well known to each other).

Technologies which can be used to replicate the salience of social interaction, or "social presence," have developed a spectrum of possible "presence levels," with higher presence occurring in environments closer to a FTF encounter. For example, text chat does not appear to resemble FTF as much as videoconferencing, due to the kinds of information available through each. The task of understanding more deeply how the perception of social presence affects language learning becomes two-fold: what specific social presence-generating elements facilitate learning a language, and how can these elements be replicated effectively in mediated environments? While much classroom-based research has explored potential applications of technology with the aim of providing a more socially rich interactive experience for language learners, unsupported assumptions are made as to what constitutes social presence and how this will lead to lasting learning gains (Lawson, Comber, Gage, & Cullum-Hanshaw, 2010). The aim of the current research is to help fill this significant gap in our current understanding of how we can use technology to provide those social elements which may be conducive for language learning, focusing on listening skills. If certain social features more readily facilitate cognitive processes involved in language learning, such as more focused attention and recall, then different technologies will be more or less appropriate as educational tools.

1.2 Research Questions

The research questions addressed are as follows:

1. What are the cognitive learning gains, as measured by error rates and reaction times, of various mediation types used for the acquisition of pitch accent? To what extent will these gains differ between audio-only (asynchronous), video (audiovisual asynchronous), and videoconferencing (audiovisual synchronous) conditions?
2. If the additional social cue of mutual eye gaze is added to a videoconferencing interaction, does this significantly increase learning gains due to higher perceived social presence?

1.3 Significance of the Current Study

There are limitations in the current research that need to be addressed in order to more fully understand how technology such as videoconferencing can effectively be applied to language learning. The focus of language learning-specific videoconferencing research on pedagogical applications to classroom tasks is evident in the research question framework used in many of these studies. The approach often taken holds a position that the technology is mostly static, and so asks how the behavior of instructors or students can be adapted to better fit the use of videoconferencing through the intervention of specialized task design. The main measure of the effectiveness of videoconferencing is how it aids completion of a particular task or affects a particular attitude, and this is often determined using students' self-reports instead of more objective measures of learning. For example, Jauregi, de Graaff, van den Bergh, and Kriz, (2012) used bi-weekly surveys of Dutch L2 students' attitude and satisfaction as the main measure of motivational increase due to videoconferencing sessions with native speaking tutors. Lee (2007) used brief reflections written after two videoconferencing

sessions by Spanish L2 students to assess their reactions to the technology's impact on the lessons. Wang and Chen (2012) used an exit survey composed of Likert scale and open-ended questions as the measure of English L2 students' sense of comfort, community, and involvement in a short online course using videoconferencing as the medium of instruction.

In a more comprehensive study, Yamada (2009) used self-reported confidence in grammatical accuracy by English L2 learners as well as the specific measure of number of self-corrections in a comparison of various mediated environments for a collaborative decision-making task in the L2. Both questionnaire responses and task session recordings were analyzed for 20 peer-to-peer pairs evenly divided among four conditions: videoconferencing, audio conferencing, text chat with a static image of the interlocutor on screen, and text chat with no image. While the study showed that less confidence in grammatical accuracy was perceived in the videoconferencing condition (which did not have any text chat capability), it also showed significantly increased self-corrections as compared to the text chat conditions, offering evidence that videoconferencing can be comparable to the social awareness of FTF language learning. The author observed that the learners in the videoconferencing condition would often try to correct their grammatical errors in response to the facial expressions of their partners, and argued that this is similar to learner behavior in FTF interactions. However, similar to the studies mentioned above, the study did not pretest or posttest the learners' language skills, such as competency in the L2 grammar.

This research is useful in establishing possible pedagogical applications of videoconferencing; aspects such as L2 learner comfort and motivation are important for

language instruction. However, the ultimate goal of language learning is improved competence in the language. Measures that more directly show improved language use, perception, and recall would provide stronger evidence that mediated pedagogical tasks can accomplish similar results as FTF pedagogical tasks – namely the retention and recall of L2 features as evidenced by improved test performance. Evidence of these cognitive effects of mediated language input and instruction is needed for a more complete understanding of the behaviors being observed.

Furthermore, while videoconferencing can be considered to be similar to a FTF interaction due to its inclusion of synchronous audiovisual information of the interlocutor's face and voice, lack of particular social behaviors such as mutual eye gaze in videoconferencing can inhibit the perceived social presence (Bondareva, Meesters, & Bouwhuis, 2006). Mutual eye gaze has been shown to have an impact on L1 language use, such as being used to signal attention to an interlocutor's speech (Argyle & Cook, 1976), as well as a lack of it leading to altered utterances (Goodwin, 1981). Research has also shown that eye gaze can impact recall of speech (Fullwood & Doherty-Sneddon, 2006) and may take priority in attention given to co-occurring social behaviors (Neureiter, Fuchsberger, Murer, & Tscheligi, 2013). The social context in which eye gaze occurs during speech can also be impacted by cultural experience and habits (Gumperz, 1982, 2003; Levinson, 2003).

Within L2 learning, there is evidence that mutual eye gaze between an L2 learner and a tutor can reliably predict the learner's successful correction of mistakes. McDonough, Crowther, Kielstra and Trofimovich (2015) found that the duration of an L2 learner's eye gaze at the tutor and duration of mutual eye gaze between both the learner

and the tutor were significant predictors of certain learner responses. Specifically, in response to a grammatical recast by the tutor, learners were more likely to respond with a new utterance that repaired the initial error, as opposed to repeating the initial error or giving no response at all. In particular, mutual eye gaze increased the odds of a successful correction by the L2 learner from 9% to 28%, in a FTF environment. However, it is unknown if this effect would occur in a similar way through mediation. Evidence of the effects of mutual eye gaze in a mediated environment is also needed to better understand its usefulness as a potential L2 learning tool.

If eye gaze in videoconferencing was corrected to be more direct and perceived as mutual, that may enable the mediation environment to more effectively encourage the eye gaze behaviors of a FTF interaction, and thereby replicate the effects of FTF learning contexts. The current study specifically explores which technology may best recreate the cognitive FTF L2 learning experience, which is necessary to create mediated L2 learning tools and apply them effectively. This could further help to provide a viable alternative to students without access to true FTF interactions with native or near-native speakers.

1.4 Conclusion

This chapter introduced some of the foundational issues of learning an L2 with the aid of technology, and specific features of learning Japanese as an L2 that can be explored to further our understanding of some technologies' potential to assist that learning process. Research questions designed to investigate these issues were introduced, and the answers to these questions can inform on which mediation types may be most practically applied with the greatest benefit to L2 learners.

Chapter 2 critically reviews literature relevant to the current study, including social presence in mediated L2 learning, eye gaze as a social contextualization cue, and cognitive linguistic accommodation, as well as discussing the relation these areas have to the current investigation. Chapter 3 provides an overview as well as detailed information about the current study's procedures, instruments, and participants. Chapter 4 presents both data analysis methods and the results of the current study, and Chapter 5 interprets and discusses these findings, looking at possible implications to mediated L2 learning and further study.

Chapter 2: Literature Review

2.1 Introduction

This chapter reviews three areas of literature which provide a context for the current research. Section 2.2 begins with a brief introduction of the construct of social presence, and then focuses on reviewing the literature relevant to mediated adult L2 learning, providing related background information from child studies as well as L1 studies. Section 2.3 focuses specifically on eye gaze as a social contextualization cue, providing evidence that it functions as a cognitively important element of social presence, and therefore may impact L2 learning and accommodation in mediated environments. Section 2.4 discusses cognitive linguistic accommodation, a phenomenon which is both automatic in L1 and L2 language, and can be socially impacted. This section provides further evidence that attention to the type of mediation used when teaching adults an L2 (specifically an L2 sound system) is needed to better understand the impact of social presence and interaction in mediated lesson environments. Finally, section 2.5 bridges together these areas of literature in a discussion of how they have influenced the design of the current study.

2.2 Social Presence in Mediated L2 Learning

Social presence is the term used to refer to the construct encompassing the focus and awareness one has of their relationship with others in an interaction, and is used often when discussing how individuals perceive a social interaction in mediated communication. Short, Williams, and Christie (1976) established social presence as the "degree of salience of the other person in the interaction and the consequent salience of the interpersonal relationships" (p. 65). They argued that qualities such as intimacy and

immediacy are impacted by the degree of social presence a medium can allow. In this context, intimacy refers to the perceived quality of the individuals' understanding of each other, while immediacy refers to the psychological distance (or perception of physical closeness) communicators feel to each other (Guichon & Cohen, 2014). Short et al. (1976) also argued that "it [social presence] varies between different media, it affects the nature of the interaction and it interacts with the purpose of the interaction" (p. 65). When considering different media for such an interactive process as language learning, social presence is something that impacts the pedagogical choices made.

Social presence has been shown to influence language learning in children in both L1 and L2 learning situations, such as learning a minority language in bilingual households (De Houwer, 2007; Pearson, 2007). Language learning through mediation such as video recordings of child-directed speech has proven ineffective for 9 month old infants learning L2 phonemes (Kuhl, Tsao, & Liu 2003), as well as 18 month old infants learning new L1 vocabulary (DeLoache et al., 2010). In both sets of experiments however, the FTF condition did result in learning.

FTF interaction is composed of a set of elements that occur together: access to both visual and auditory information for speech, synchronicity of interaction with no artificially imposed delays, the potential for direct eye gaze (or eye contact), as well as other information like bodily gestures and affect. All of these elements can occur when interlocutors are physically co-present in the same location under normal FTF circumstances, and can contribute to the perception of a social or emotional link. Individual elements are easily lost, however, when mediation is used to compensate for lack of co-presence. Audio-only CDs do not have visual information such as lip

movements and facial expressions. Prerecorded video provides audiovisual speech, but is not synchronous, so there is less pressure than a real-time encounter which would require quick understanding in order to follow utterances that occur only once, as well as requiring responses for the interlocutor to move on – in video recordings, the speaker continues regardless of a response given or not given. This means that while social cues are important for language learning even at the earliest stages of acquisition, mediation may not be effective for early language learners.

For adult L2 learners who are already well-established in their L1, audiovisual and computer mediation shows more promise as an effective medium for the social information needed to learn and use an L2. For example, Arnold and Hill (2001) tested adult English L1 French L2 speakers' abilities at comprehending both L2 input (spoken passages in French) and phonemically difficult auditory L1 input (Scottish-accented English, and semantically and syntactically complex Standard English). The participants were at an intermediate level for their L2 of French. They were tested in both audio-only and audiovisual conditions, in order to investigate any advantage the addition of visual information may provide when trying to comprehend either a developing L2 or a native L1. The authors measured the participants' comprehension of the spoken passages instead of shadowing (repeating words), asserting that this is a more reliable measure of enhancement by visual input. This task was also a mediated language presentation that more closely resembled a social situation than shadowing. Significantly improved comprehension was observed with all three passage types when the visual information was included. Research has also shown that highly proficient L2 speakers perform significantly better when provided with visual cues (facial displays and lip movements),

as opposed to audio-only mediation (Navarra & Soto-Faraco, 2007). This is consistent with findings which indicate that for adult L1 speakers, perception of language integrates audio (phonemic) and visual (lip movement) information to a high degree and may use visual information even more when the audio signal is difficult to comprehend or conflicts with the visual signal (at least in speakers of English and French Canadian; see Dupont, Aubin, & Menard, 2005; Massaro, Thompson, Barron, & Laren, 1986; McGurk & MacDonald, 1976).

Another feature of social language use is that of synchronicity – when speaking to another person, the interaction occurs in real time. This is likely to impact the immediacy individuals feel when communicating. A now widely available and accessible technology that provides synchronous audiovisual information is videoconferencing. In the field of Computer Assisted Language Learning (CALL), there has been some research into the use of videoconferencing specifically for its capacity to bring mediated social interaction into the classroom as a way to provide access to native speaker input when it would otherwise be unavailable. Generally, these investigations have found that students respond positively to videoconferencing when it is a one-to-one or small group environment, and feel the ability to see the instructor or another student is helpful to their learning (e.g., Jauregi et al. 2012; Lee, 2007; Yamada, 2009; Yamada & Akahori, 2007).

Social presence in an educational setting is typically assumed to be important, as it is seen to support student learning through placing emphasis on immediate and effective (or intimate) communication, which can situate what is said and heard within a student's more general knowledge and facilitate a connection there (Guichon & Cohen, 2014; Zhan & Mei, 2013). Studies have investigated social presence effects in mediated

learning environments mostly through the use of student self-reports of their impressions of their own language skill gains, as well as their thoughts on the social elements that are included in the environment. Lee (2007) explored the effects of a "collaborative and non-threatening learning environment" (p. 638) in which expert speakers of Spanish were paired with Spanish learning students to perform two jigsaw-like tasks via videoconferences. The author found that the students had a positive reaction to the videoconferences, as it raised awareness of the importance of proper pronunciation, and exposed students to more vocabulary and different dialects. However, communication was not felt to be "complete" because even though they could see each other, most participants did not use body gestures during their interactions and commented on how they often forgot to look at the camera directly to make eye contact.

Similarly, Jauregi et al. (2012) aimed to determine if videoconferencing with native speakers could increase motivation for foreign language learning students, and also if the students' proficiency levels affect this possible increase. To investigate, Czech university students of Dutch language, at two different levels of proficiency (beginner and intermediate), were paired with Dutch student-teachers. The pairs used videoconferencing to interact three times per week, for 10 weeks, and surveys were used to assess the students' attitudes and satisfaction bi-weekly. While the results showed the students felt positive towards the sessions overall, only the beginner-level students had significant changes in "feeling more competent" in the language. However, motivational attitudes may not have improved due the use of the technology alone, as the student-teachers purposefully used "motivational techniques" in their teaching during the videoconferencing sessions. Also, there were signs that some of the perceived skill gains

could be due to novelty effects. Students in the intermediate language group received prior exposure through participating in the same experimental condition in a pilot study. The experience was novel for the beginning students, which could have also impacted their survey ratings.

Some studies have focused on both qualitative and quantitative effects. Yamada and Akahori (2007) examined students' perceptions of social presence by comparing different technologies (videoconferencing, audio conferencing, text chat with static interlocutor image, and text chat without image), and analyzed language production (e.g., number of turns, number of grammatical errors during the interactions, number of self-corrections). Assigned to one of the four conditions and placed in separate rooms, randomly assigned pairs of students used laptops to have 15 minute discussions on an assigned topic (selecting one of four possible school teachers) in which they tried to use target formulaic expressions in the L2 that were provided in on-screen lesson material. It was found that perception of interlocutor presence was greater when an image was available. Analysis of language production data revealed a main effect of interlocutor image: the presence of an image resulted in more natural-language utterances. On the other hand, the text-based conditions yielded fewer grammatical and lexical errors.

Similarly, Yanguas (2010) also compared different levels of mediation, including a non-mediated FTF condition. Fifteen pairs of Spanish L2 university students were assigned to either a videoconferencing, audio-only, or FTF condition, and then were observed completing the same jigsaw task in their L2. This task required the learner pairs to plan for a hypothetical backpacking trip across Latin America, using a limited set budget and selecting from a pool of possible items. The partners each received a different

set of eight item pictures with prices (one half of the 16 total possible items), and together had to select only four items from each set to take, without exceeding the budget. In a qualitative analysis, it was found that both the videoconferencing and audio-only groups were comparable to FTF with respect to amount and quality of negotiations for meaning, but videoconferencing was more similar to FTF with relation to gestures used to convey lexical items unknown in the L2.

Such research shows that learning an L2 through mediated audio-only input is possible for adult L2 learners, but when too many social elements are lost (or as the input becomes more mediated and thus more removed from a FTF situation) auditory input becomes harder to understand. Though it was not the specific focus of the above research, this indicates that not just any method of mediation may be helpful to the learning of L2 sound systems. While interest in the possibility of adult L2 learning that is facilitated by mediated audio and/or visual input is understandably high in L2 pedagogy (Ducate & Arnold, 2011), due to the fact that classroom time is so limited, work on the cognitive effects of social elements should be very informative for determining which technologies may be good candidates for mediated learning of phonology. The indication is that technologies that allow for both audio and visual information together are likely to be more effective.

2.3 Eye Gaze as a Social Contextualization Cue

Individual elements may be more necessary than others to enhance the cognitive perception of social presence in a mediated language learning environment. Bohannon, Herbert, Pelz, and Rantanen (2013) observe a general trend in technology and behavioral science research showing that more natural eye gaze in videoconferencing leads to richer

and more efficient communication (or communication which has a reduced potential for uncertainty and equivocation). While steps can be taken to improve the impression of eye gaze in widely available videoconferencing systems, social presence is more noticeable when true eye gaze is achieved (Doherty-Sneddon et al., 1997; Bondareva et al., 2006). This can be accomplished through physical camera placement or using multiple cameras, such as how Neureiter et al. (2013) placed cameras both within and around the monitor screen. Similar effects can also be achieved through software, such as the image-correcting computer algorithm employed by Yip and Jin (2003), which rotates the image of the eyes and the light glare within them by a few degrees to reorient the apparent gaze direction.

There is evidence that eye gaze as a social behavior may take cognitive precedence over other potential behaviors. In a study that explored the interaction of eye gaze and gestures, Neureiter et al. (2013) added cameras to a traditional videoconferencing setup in order to control eye gaze direction and to show the participants' hands and arms on screen, providing gesture information. Cameras were either placed above the screen (so that the participants appeared to be looking down, and not making eye contact) or within the screen (so that participants would be looking directly at the camera, making eye contact). In the condition which allowed direct eye gaze, participants perceived more social presence (as measured by a self-reporting questionnaire) and took notice of gestures, as opposed to the condition without direct eye gaze. Access to gestures is another way in which videoconferencing can more closely resemble a FTF interaction, but it would appear that cognitively, eye gaze is given priority, and once it is established, additional elements may come into play.

Eye gaze also has an impact on cognitive processes needed for sustained learning, such as recall. Fullwood and Doherty-Sneddon (2006) investigated the impact of eye gaze over a videoconferencing link on memory recall, which they argue is often an indicator of perceived social presence. Participants were shown pre-recorded videos, but told they were seeing real-time presentations over a videoconferencing system, and so from their perspective the presentations were synchronous. One condition had direct eye gaze (the presenter looked directly at the camera for 30% of the recording, at pre-determined points throughout) and one did not (the presenter looked at the monitor screen below the camera). Even though both presentations were matched for variables like length and content, participants could more accurately recall the information from the presentation with direct eye gaze. A second experiment in which only the audio portion of the presentations was heard showed no difference in recall, but did yield similar recall rates to the direct eye gaze condition of the first experiment. The authors assert that this shows that the lack of eye gaze had a negative impact on perceived social presence, leading to less recall. While the authors associated the participants' learning with the social presence generated by each condition, they argue that lack of eye gaze may have been perceived as gaze aversion, or purposeful lack of eye contact, which could lead to interlocutors forming negative attitudes towards the presenter that then adversely affect any social presence perceived. The common finding of these studies is that the social behavior of eye gaze may be key in helping interlocutors cognitively perceive more positive FTF-like conditions.

Eye gaze can be viewed as a nonverbal contextualization cue, or a paralinguistic cue that is deeply socially embedded and serves to prime habitually formed cognitive

perceptions of an interlocutor's intended pragmatic message, in conjunction with the lexical material of that message (Gumperz, 1982, 2003; Levinson, 2003). For example, the use of a particular intonation can provide further meaning to an utterance (as in sarcasm), or a particular body position paired with an utterance can change the interpretation of that utterance. Such cues are culture-specific. Contextualization cues will automatically activate through a mostly unconscious process that then relays the social content, or culturally habitual pragmatic patterns, of an interaction. In some cultures, for example, using direct eye gaze when saying something provides the further context of social dominance in conjunction with the lexical message. If this pattern is observed frequently over time by a member of the culture, such as a child raised in that culture, the occurrence of direct eye gaze will automatically prime the interpretation that the utterance is being said in the context of a socially dominant position held by the speaker. This social content functions as a prime for possible perceptions of the interlocutor's intended message as influenced by cultural experience (Levinson, 2003).

Because research from various fields indicates that eye gaze functions as a social behavior that is closely tied to social presence in verbal communication at a fundamental cognitive level, investigation of this particular aspect in mediated language learning is warranted. Furthermore, this investigation is needed to clarify the impact of eye gaze on speech perception. The studies described above have not explicitly focused on speech perception as an area that may be impacted by direct eye gaze, and it is possible that such an impact could be positive (such as increasing social presence) or negative (such as drawing attention away from the mouth, making visual articulation cues less salient).

2.4 Cognitive Linguistic Accommodation

Phonetic accommodation is a cognitive phenomenon that has been gaining attention over the past two decades. A good description of the essence of this effect is when "a listener…adopts some aspects of an interlocutor's acoustic-phonetic repertoire, engaging in phonetic convergence during conversational interaction" (Pardo, Jordan, Mallari, Scanlon, & Lewandoski, 2013, p. 183). This description indicates that phonetic accommodation is in reality a set of subtle effects due to cognitive processes that can occur simultaneously and result in adaptation to an interlocutor's speech style. The phonetic aspects involved are converged and perceived at a level beyond a simple shifted representation of the sounds. The aspects of speech which have accommodated most noticeably are the fundamental frequency (e.g., Gregory, Green, Carrothers, Dagan, & Webster, 2001; Winters & Grantham O'Brien, 2013), vowel spectra and durations (e.g., Babel, 2010; Lelong & Bailly, 2011; Pardo, Gibbons, Suppes, & Krauss, 2012; Winters & Grantham O'Brien, 2013), and voice onset times (e.g., Nielson, 2011; Sancier & Fowler, 1997).

Accommodation effects of this kind are initially automatically activated (e.g., Pickering & Garrod, 2004; Trudgill, 2008), and can occur for both same-language and different-language speakers. The automaticity of accommodation is reflected in its occurrence in a range of communicative settings, from non-interactive lab tasks in which monolingual participants listen to recorded L1 speech (e.g., Babel & Bulatov, 2011; Namy, Nygaard, & Sauerteig, 2002), to more natural interactive L1 speech (e.g., Pardo, 2006), to L2 learners with extended ambient exposure to either an L1 or L2 (e.g., Sancier & Fowler, 1997). What the literature also shows is that the automatic processes of

phonetic accommodation can be impacted by social circumstances such as familiarity with the interlocutor (Lelong & Bailly, 2011) and social bias towards differing L1 dialects (Babel, 2010), which have resulted in complex individual differences (Pardo et al., 2013).

Pardo (2006) found that gender and social status effects are highly variable and somewhat contradictory throughout the literature, and that the type of task may be the more important factor in how much and how long accommodation effects impact individuals. In her study, pairs of unacquainted participants sat on either side of a divider so that they could not see each other, but could hear each other. Both individuals had a set of maps, and one participant was tasked with giving the other directions on how to mark the right path on the map to a set of five destinations. The participants were allowed to talk freely to complete the task, and tokens of particular words were captured from both during the task. Tokens of the target words were also recorded both before and after the map task by having each participant read a list of words aloud. These tokens were then rated for accommodation by another group of participants. It was found that accommodation increased over time during the map task and continued into the post-task list-reading session. While no specific duration is reported, the accommodation showed persistence after the interaction ended, which is longer than accommodation from listening to prerecorded words in isolation. Pardo (2006) concludes that tasks that are interactive and maintain high levels of attention, such as the map task, are likely to produce more reliable and persistent accommodation overall. This indicates that social presence plays a significant role in the process. The fact that the phenomenon is wide-ranging in regards to speech features which can be both automatically activated and

socially impacted makes it a relevant factor to consider when investigating the social presence generated by technology in L2 learning situations.

The process of phonetic accommodation can be beneficial to furthering the progress of L2 learners in the areas of speaking and listening skills (Trofimovich, 2013). For L2 learning, even short exposures to the speech of native speakers (or advanced non-native speakers) of an L2 is enough to trigger accommodation. Maintained accommodation effects in an L1 can build over time and can persist to subsequent tasks; Pardo (2006) found phonetic convergence in tokens recorded during the post-task session which followed an interactive map task. L1 accommodation effects can also generalize to similar phonetic features (Nielson, 2001). Concurrently, Wang (2001) found that L2 learners maintained improved ability to perceive Mandarin tone differences 6 months after eight 40 minute listening exposures given over two weeks. Taken together, this suggests that in situations where long-term exposure to models of the L2 is limited, many short-term exposures over time can still be beneficial. Facilitating such short-term exposures may be more easily managed than long-term ones, and the use of mediation such as videoconferencing would be ideal for these exposures if cognitive perception of social presence is indeed similar enough to FTF interactions. There is currently a need for research on accommodation effects through videoconferencing mediation to investigate this possibility, however, as this has not been a focus of recent studies.

2.5 Conclusion

Given the important studies reviewed above, the foundation laid impacts and shapes the current study in several ways. The current study is similar to Yamada and Akahori (2007) and Yanguas (2010) in that various levels of mediated L2 lessons are

compared to each other. While textual elements were considered in the prior work (such as text chat), the focus here is audio and visual information as it relates to the ability to perceive and produce Japanese pitch accent. Some studies have noted the need to incorporate language assessment measures that show student progression in specific skill competencies to provide a more complete understanding of the perceptions learners have of their own skills (e.g., Bilbatua, Saito, & Bissoonauth-Bedford, 2012; Jauregi et al., 2012), and this is the approach of this research. In order to further investigate social presence effects on cognitive learning gains, quantitative measures of behavioral data are used in the form of a pretest/posttest/delayed posttest design.

Using this design, the current study also further investigates potential cognitive effects of indirect eye gaze (currently the standard in videoconferencing) and direct eye gaze on speech perception. Through comparison of cognitively measured learning gains in videoconferencing conditions that either have eye gaze which has been corrected to be direct or not, potential effects of direct eye gaze on perception of social presence and language learning in mediated environments can be observed. The addition of direct eye gaze is expected to be beneficial due to previous research indicating both its prioritization over other cues and impact on cognitive processes in mediated audiovisual communication.

In order to gather more information to aid in understanding any potential effects, two surveys are also used. The first is a demographic survey designed to look for relationships or confounds that may impact the results of the testing tasks, such as exposure to tonal languages and familiarity with the technologies being used. The second is an exit survey which primarily explores participant motivation in completing the study

training and testing, by probing to what degree that motivation is intrinsic or extrinsic. This survey also asks about how comfortable participants are being viewed by a tutor as they work through lessons, and their likelihood to further study the Japanese language.

Finally, the current study also investigates the possibility of accommodation effects through videoconferencing mediation by using a design that activates phonetic accommodation through brief lessons in an L2. While accommodation can be activated in highly mediated tasks such as repeating pre-recorded single words, a language lesson setup is more conducive to also stimulating the perception of social presence, as this directs attention towards the goal of learning how to communicate with L2 speakers and gives the language being used a social purpose. A cognitive measure of improvement in listening skills for the L2 (collected error rates) allows for comparison of mediation levels. Reaction times are collected to help detect any potential speed/accuracy trade-off effect. Accommodation will likely occur in all levels, but the cognitive measures will allow for observation of differences in the strength of accommodation, as reflected in actual learning gains.

Chapter 3: The Current Study

3.1 Overview

The current study investigated the first research question by direct comparisons of different levels of mediation in L2 learning lessons, using cognitive measures of actual learning gains. The second research question was investigated by comparing the cognitive learning gains of two videoconferencing conditions, in which the manipulated variable was direct or indirect eye gaze allowance between the video feeds of both the student and tutor.

Five sets of data were collected and analyzed over the course of the study during two sessions for each participant (see Table 3.1).

Table 3.1

Overview of Study Design

Session	Part	Instrument				Data collected
1	1	Survey 1				Demographic
	2	Pretest				Error rates; reaction times
	3	Training Conditions: audio-only	video	videocon-ferencing	videocon-ferencing with eye contact	Lesson item answers; verbal production (not analyzed in current study)
	4	Posttest				Error rates; reaction times
2	5	Delayed Posttest				Error rates; reaction times
	6	Survey 2 (Exit Survey)				Motivational subscales; comfort with tutor presence; future study

What follows is detailed descriptions of the participants, materials, and procedures of the current study. Each of the instruments, including the four training conditions, and the data collected by each are further explained below in sections 3.2.4 and 3.2.5.

3.2 Methods

3.2.1 Setting. The experiment was conducted in a setting that is similar to the setting in which language learners would typically engage in computer-assisted language-learning tasks: they were seated in front of a computer monitor in a quiet room.

3.2.2 Participants. English L1 participants with no prior educational study of Japanese were tested. Further, participants were not included if they had studied Chinese, Korean, or Thai. Although these are not pitch accent languages, their similarity to Japanese either phonemically or with regard to pitch changes (such as the tonal minimal pairs of Chinese) could result in L2 speakers of these languages being more sensitive to the minimal pairs used in the study (Wayland & Li, 2008).

Participants were both male (N = 37) and female (N = 45), and were balanced within each of the four conditions. The exact counts for each condition was: Audio, male = eight, female = 14; Video, male = six, female = 15; Videoconferencing without eye contact, male = 11, female = eight; Videoconferencing with eye contact, male = 12, female = eight. Gender differences have been seen in the literature on accommodation, but the reported differences are highly variable and contradictory between studies, and it has been argued that it is more likely task differences which are responsible for the effects (Babel & Bulatov, 2011; Namy et al., 2002; Pardo, 2006). Four participants were removed for not meeting the experiment requirements, such as completion of all tasks and tests.

All participants were college attending adults, with a mean age of 20 (range = 18-42), to ensure that they had an established L1 and were similar to most second language students who may encounter mediated environments for the purposes of language learning. They were randomly divided across the four conditions, to create groups of comparable size (see above). All participants received course credit for their participation.

3.2.3 The learning task. The task focused on an aspect that L2 learners typically have trouble with, namely perceiving sound system differences between the L1 and L2. Past studies have addressed this most effectively by targeting an L2 which exhibits a feature that is not present in the L1, or that is present but does not produce minimal pairs in the L1 (Kuhl et al., 2003; Navarra & Soto-Faraco, 2007). Here, a pitch accent placement difference in Japanese minimal pairs was used due to the fact that non-native speakers do not typically notice it unless attention is brought to it (Wayland & Li, 2008). This type of feature is noticed on a more subconscious cognitive level of processing (as discussed in Chapter 2 with accommodation), but can be trained in non-native speakers through awareness building (Sugiyama, 2006; Wayland & Li, 2008). The subtlety of this feature allowed for the apparent focus of the lessons to be on more salient units, such as words and sentences.

3.2.4 Materials. Materials consisted of the testing materials and the lessons, as well as two surveys.

3.2.4.1 Surveys.

Survey 1. The first survey was a brief demographic questionnaire administered after informed consent and before pretesting, which asked about:

- age
- sex
- ethnicity
- handedness
- vision or hearing problems
- musical training
- languages spoken
- frequency of use of those languages
- extent and type of exposure to other languages
- familiarity with language learning software such as CD's and DVD's
- familiarity with videoconferencing technology

Age, sex, and ethnicity were collected as basic demographic information, in order to have accurate and more complete descriptive information about the participants. Over the past two decades, some patterns between handedness and language learning and use have been observed (Boiteau, Smith, & Almor, 2017). Handedness information was collected in order to explore any potential trends between handedness and learning to recognize pitch placement. Vison and hearing problem information was collected in order to assess whether participants may have difficulty with the lesson and testing materials, as these materials had visual and audio components. Information on what languages participants speak, how often, and what languages they are exposed to was collected to ensure that their language use did not give them an unfair advantage in the study, and to explore any potential confounds or trends related to other languages they may speak or hear regularly. Familiarity with language learning software and with videoconferencing technology was also collected to explore for any potential confounds or trends concerning frequency of use of these technologies and the study results, as the current study made use of these technologies. Finally, information on musical training was collected because this particular type of training focuses attention on sounds, tones, and pitch. Extended musical

training was seen as potentially having a relationship to the speed and accuracy with which one may learn to recognize minimal pairs in a pitch accent language such as Japanese, and this information was collected to explore that possibility within the parameters of this study.

Survey 2. The second survey was administered after the delayed posttest, and asked about the participants' motivation to complete the lessons and testing materials, as well as comfort in being observed by a language tutor when working through lessons, using 7-point Likert scale questions.

The survey items in the first part identified participants' situational intrinsic motivation (or motivation to complete tasks for the pleasure and satisfaction they provide) and extrinsic motivation (or motivation to complete tasks through some sense of external obligation). This was explored because of its potential to further clarify participant reactions to various types of technology and the perceived social presence each affords. Deci and Ryan (1985) proposed self-determination theory as a way to understand the quality and nature of an individual's motivation to perform a task. They developed a view of self-determination, or how much someone feels they have freely chosen to do something, as a continuum ranging from intrinsic motivation to extrinsic motivation. Intrinsic motivation reflects higher self-determination because the task is perceived as inherently pleasing in and of itself. Extrinsic motivation is lower in self-determination because the task is perceived more as a means to an end; even if the means were chosen, they are chosen because of the desired end goal, and so there is some restriction on the choice. For example, one may choose to study a subject because they feel pleasure in learning something new, or they may choose to study to obtain a high

grade in a class. While both choices are determined by the individual, the latter is being made less freely because the true goal is the high grade; this limits the range of choices to those that can lead to the goal, and the goal is more desirable than the chosen activity. This shifts the motivation for performing the task from the task itself to the goal it leads to, which is an external source of motivation. It may be possible that the type of motivation an individual has when using a meditated environment to learn a second language can impact their learning performance (e.g., Morton & Jack, 2010). Therefore, this information was seen as potentially informative to the analyses of learning gains between mediation types, especially if different motivational reactions are observed for different conditions.

This part of the survey was adapted from the instrument developed and validated by Guay, Vallerand, and Blanchard (2000). The instrument, the Situational Motivation Scale (SIMS), was developed and tested to assess four subscales, as suggested by Deci and Ryan (1985) in their self-determination continuum, of potential situational motivation experienced by participants when completing a task. These four subscales, as described by Guay et al., include intrinsic motivation, identified regulation, external regulation, and amotivation. Intrinsic motivation (described above) is reflected in agreement with statements such as "I completed the lessons because they are fun." Identified regulation is seen when behaviors are perceived as self-chosen and valuable, but performed to accomplish an external goal, and thus extrinsically motivated. An example of identified regulation would be the statement "I completed the lessons because I believe they are important for my advancement." External regulation is seen through behaviors performed to either avoid negative consequences or to receive rewards, and so

are also extrinsically motivated. An example of this would be the statement "I completed the lessons because they are required to pass the class." Amotivation is reflected through behaviors that are perceived as having no real connection or impact on the outcomes; there is no sense of purpose or possibility for change due to the behavior being done. An example would be the statement "I don't know why I completed the lessons – I'm not sure they have any worth."

The second part of the survey asked participants to report their general comfort or anxiety when they are observed by a tutor during a lesson, as well as their preference for working with a tutor either present or not present during an assignment. The final two items of the survey asked participants if they would consider studying Japanese after the completion of their participation, either through self-study or through formal education (see Appendix A for both surveys).

3.2.4.2 Test materials. The testing materials were used for the pre, post, and delayed posttest. These tests were a same/different discrimination task and a picture recognition task, administered using DMDX, a Win 32-based display system capable of measuring responses and reaction times to visual and auditory stimuli (Forster & Forster, 2003).

Same/different discrimination task. Wayland and Li (2008) have shown that same/different discrimination tasks are effective for measuring improvements (from pretest to posttest) in non-native speaker recognition of tone contrasts in an L2. This type of task can also be used to train for better recognition, but this was observed when the task gave feedback to the participants (Wayland & Li, 2008). As this task in the current study did not give feedback on correctness, no training effects due to the pre/post testing

were expected, and any training effects presumably occurred equally in all participant groups. The same/different discrimination task in each test used the same 22 test items; 11 "match pairs" and 11 "contrast pairs." In a match pair item, two different tokens of the same Japanese word were presented, while in a contrast pair item, one token each of two different Japanese words were presented, with only the pitch pattern being the difference (see Table 3.2; see Appendix B for a full list of all pairs).

Table 3.2

Same/Different Discrimination Task Test Item Examples

Word pair type	Token one	ISI	Token two
Match	「箸」[ha*ʃi] "chopsticks"	500ms	「箸」[ha*ʃi] "chopsticks"
Contrast	「箸」[ha*ʃi] "chopsticks"	500ms	「橋」[haʃi*] "bridge"

Note. Japanese characters included here to help illustrate that tokens are different words. A "*" indicates the syllable with a higher pitch. ISI = inter-stimulus interval. ms = milliseconds.

Note that all word presentations for this task were audio-only. A "+" appeared on the screen while the two tokens played, one after the other, with a 500 millisecond (ms) inter-stimulus interval (ISI) between. At the offset of the second token, "same?" appeared on the screen, and participants responded by pressing a key to indicate that the pair was either the same word twice or two different words. While times shorter than a 500ms ISI (such as 250ms) are likely to be long enough for phonetic processing, this ISI provided a sufficient enough delay to ensure that the participant's comparison between the two words included processing of phonetic differences such as those present in tonal languages (Wayland & Li, 2008). Wayland and Li (2008) compared English L1 speakers'

ability to discriminate between Thai tones at an ISI of 500ms and 1500ms, and found that performance was comparable at both ISIs, while the ISI of 500ms had an increased demand on working memory load. The 500ms ISI provided enough processing time for English L1 speakers to be able to discriminate, but also maintained enough difficulty so that improvement in discrimination would show variability (Wayland & Li, 2008).

The stimulus tokens in the testing materials were disyllabic (as the example above shows), and the feature was approached as a lexical one to maintain the most controlled conditions possible. Additionally, as Sugiyama (2006) found that pitch placement differences in F0 are most robust when sentence medial, all tokens were digitally removed from full sentence utterances produced by a native speaker in which the token occurred in a sentence medial position and the speaker pronounced each word clearly. The tokens were removed carefully, to ensure that the phonemes were clearly recognizable but not overlapped by the preceding or following phonemes. This ensured more distinct and reliable F0 contrasts with regards to the minimal pitch differences.

Picture recognition. The picture recognition task in each test used 18 recognition items which were also used in the same/different discrimination task; nine "correct" and nine "incorrect" items. In each "correct" item, a picture of an object was presented, followed by the auditory presentation of a token which correctly identified that object. In each "incorrect" item, a picture was presented and followed by a token which did not identify the pictured item, as it was the other member of the minimal pair (see Appendix B). When test items were presented, the picture appeared for 2000ms. Following this, there was a 200ms ISI, after which a "+" appeared on the screen while a token was played. At the offset of the token, "correct?" appeared on the screen, and participants

responded by pressing a key to indicate that the word they heard correctly identified the item in the picture or not. As with the same/different discrimination task, no feedback on correctness was displayed for the participants during the task, so no training effects due to the task alone were expected.

The order of item presentation was automatically randomized for each test by the DMDX software. All sound files of these words are recordings of a native speaking "tutor" saying the words clearly, with two tokens of each word.

3.2.4.3 Lesson materials. The lesson materials were two brief (approximately 15 minutes in length) beginning-level lessons of Japanese, each using a worksheet. The worksheet for the first lesson was two single-sided pages. The first page had lesson instructions printed at the top which were read aloud by the tutor as the participants read along. The rest of this first page introduced a vocabulary list, which was displayed in a nine row, three column grid. Each vocabulary item featured a representative greyscale picture of the word's meaning, the Japanese pronunciation of the word written in roman characters, and the English translation of the word. Some items also included a further brief description, for clarification. For example, "sickle" was followed by "(farm tool)" to further clarify the vocabulary word's meaning. The participants repeated each word aloud following the tutor's example; the tutor paced the reading so that each word was clearly read and not rushed. This list included nine minimal pairs of nouns (18 items), along with seven non-minimal pair items, for a total of 25 items. The non-minimal pair items were adjectives and appeared in the example sets of the following lesson. The second page had a brief "listen and identify" exercise. The printed instructions were again read aloud by the tutor as the participants read along. This was followed by six numbered blank lines,

where participants heard a word read aloud and then had to write the word using roman characters followed by its English meaning for each of the six question items.

The second lesson worksheet was a single page which introduced simple Japanese sentence structure. The printed lesson instructions described how to form a simple sentence structure in Japanese: how to describe a noun with an adjective. These instructions included four example sentences, written both in Japanese pronunciation (using roman characters) and their English translations. The instructions were read aloud to the participants as in lesson one, and they read along. The example sentences were also read aloud, and the participants repeated each one aloud after it was read to them. This was followed by a "complete the sentence" exercise, where each question item featured an incomplete sentence written in Japanese pronunciation (in roman characters) and a blank line indicating the missing word (either a noun or an adjective). No English translation was given for the question items. The participants completed these sentences with previously presented words from the lesson one worksheet that would make the sentence semantically comprehensible (eight items total; see example below). Once the sentences were complete, the participants then read them aloud (see Appendix C for all lesson worksheets).

Examples provided: /kame wa mesu desu/ 'The turtle is female'
/hashi wa chiisai desu/ 'The bridge is small'

Example exercise items: /mesu wa _____desu/
/_____wa nagai desu/

3.2.5 Procedures and data collection. Participants met with the investigator two separate times, with approximately 5-7 days between visits, depending on each participant's availability. The first meeting consisted of the demographic survey, the

pretest, the two lessons (training), and the posttest. The second meeting consisted of the delayed posttest, the exit survey, and an oral debriefing. Participants received all lessons and testing individually. The lessons were audiotaped via cassette recorder. The total procedure lasted approximately 1-2 weeks for each participant. Procedures for collecting the data were as follows:

(1) Demographic survey. After having indicated informed consent through reading and signing a consent form, each participant read and answered the brief demographic questionnaire (see Materials, Survey 1).

(2) Pretest. Each participant was then given the pretest consisting of both the same/different auditory discrimination task and the picture recognition task (see Materials). The same/different task occurred first, followed by the picture recognition task (after participants indicated they were ready to move on to the next task by pressing a key). The pretest was administered using DMDX, which recorded the error rate and reaction time (in ms) for each item in both test tasks. All participants sat approximately forty centimeters from the computer screen. Participants wore headphones to ensure a clear auditory signal. The pretest provided on-screen instructions at the start of each task stating that the participants will either listen to pairs of words and then respond if the words are the same or different, or view pictures and listen to words and then respond if they are correctly matched using the left and right shift keys. Small stickers were placed on the keys to remind participants throughout the task of the response keys should they have forgotten. The on-screen instructions for the same/different task also informed the participants that they would first complete a "practice" trial, and that this trial would repeat if they did not get three out of five responses right. This was included to provide a

brief training to the response keys, and to ensure that test errors were not due to misunderstood directions. All practice trial items were different from the test items, and while similar in sound, were words which differed in more ways than just pitch placement, so that only test-response training was taking place during this phase of the pretest. Test items were presented in an order automatically randomized by the DMDX software.

(3) Training. Once the pretest was complete, participants were given instructions appropriate to their assigned condition and started the first lesson (see Materials). Descriptions of the procedures for the lessons during the first meeting for the four conditions follow.

Audio-only condition: This condition provided participants with only auditory information from the tutor. Participants in this condition were given the lesson worksheets and asked to read over the directions for lesson one while the computer was set up to provide a list of two recordings for use with the two lessons (both recordings were audio of the tutor's voice taken from videos of the tutor administering the lessons). The files were set up using Windows Media Player. The player was put in "full screen" mode and the participants saw a default music note image and the two listed recordings. The participants used headphones to ensure a clear signal. Once setup was complete, the participants first heard the instructions read aloud in the first recording, and then began following the instructions, which asked them to first repeat aloud each example to themselves as they listened. They then began working on completing the six "listen and identify" question items, where participants heard a word and then had to write the word using roman characters and its English meaning. Participants were told there would be a

tape recorder recording audio in the background as they did the lessons; this was clearly in view. Participants were told this was to check general participation while they repeated back items during the lessons. They were told that their name would not be attached to the recording in any way; the tapes were labeled with participant code numbers only. The purpose of the recordings was to ensure that the participants repeated the items aloud when prompted even though they were not talking to a person in real-time as well as to provide a means to know if they did not follow the instructions to repeat the items. This allowed for more consistent behavior data across all conditions. After the question items were complete, the participants heard the question items again to check their answers. Once this was done, the lesson ended. This was followed by the second lesson, following the same procedure with the addition of repeating the completed question items aloud to themselves.

Video condition: This condition provided participants with both auditory and visual information from the tutor. Participants in this condition were also given the lesson worksheets and asked to read over the directions for lesson one while the computer was set up to provide a video file for use with the exercises. The file was set up using Windows Media Player, in “full screen” mode. The participants used headphones to ensure a clear signal. The video file was a recording of the tutor administering the lesson, with the tutor’s full face and shoulders visible and facing the screen. Throughout the video, the tutor would look directly at the camera at points that seemed natural (i.e., while stating a new vocabulary word or question items, so that his face and mouth were clearly visible). Participants were asked to read over the directions again while the recorded tutor read them aloud. The tutor then began administering the first lesson. The tutor

pronounced all examples for the participants and asked them to repeat the examples after this modeling, providing pauses so that the participants had time to do so. The tutor then presented the question items, and provided pauses for the participants to write down responses (the information given in this condition and the other conditions was the same, as to ensure comparison). As described in the audio-only condition above, the participants were told there would be a tape recorder recording audio in the background as they did the lessons, which would be labeled with participant code numbers only. All other procedures for this condition were identical to the audio-only condition.

Videoconferencing condition: This condition provided participants with audiovisual information from the tutor and synchronous interaction. Participants in this condition were also given the lesson worksheets and asked to read over the directions of lesson one while the live tutor read them aloud via a videoconference call over the computer (the tutor was visible the entire time as a full-screen image). The tutor was instructed to look only at the participants' faces on the screen and not directly at the camera, as is standard practice when using typical videoconferencing equipment on a computer. As described in the conditions above, the participants used headphones and were told there would be a tape recorder recording audio in the background as they did the lesson, which would be labeled with participant code numbers only. The tutor then began administering the first lesson. The tutor pronounced all examples for the participants and asked them to repeat the examples after this modeling. The tutor then presented the question items. The tutor waited until the participants had completed a question item before presenting the next one. After the question items were complete, the tutor repeated the question items once more so that participants could check their

answers. The tutor then instructed the participant to continue on to the lesson two worksheet, and worked through it with the participant in the same manner as lesson one. Once this was done, the tutor thanked the participant for their participation, and the lessons and video call ended. All other procedures for this condition were identical to the video and audio-only conditions.

Videoconferencing with eye contact condition: A final condition did not change the type of mediation used, but instead replicated the videoconferencing condition procedures with the addition of the affordance of eye contact between the participant and tutor during the videoconferencing call. The synchronous and audiovisual aspects of the videoconferencing condition were necessary for the manipulated variable of eye gaze to affect social presence perception which can be measured as effects on learning gains. The only difference between the two videoconferencing conditions was the presence or absence of an apparatus for eye gaze correction, which was fixed to the computer monitor. This type of apparatus provides eye gaze correction in a more natural and reliable way than currently available image rendering programs. The apparatus used, the ProPrompter Desktop, follows the basic design principles for eye gaze correction first used by Randall Smith and William Newman in their "video tunnel" (reported in Buxton & Moran, 1990). It uses both full and half silvered mirrors positioned in front of the computer screen and webcam, which redirects the on-screen image to correct the line-of-sight (see Figure 3.1). The image of the tutor was smaller than full screen, in order to work with the apparatus.

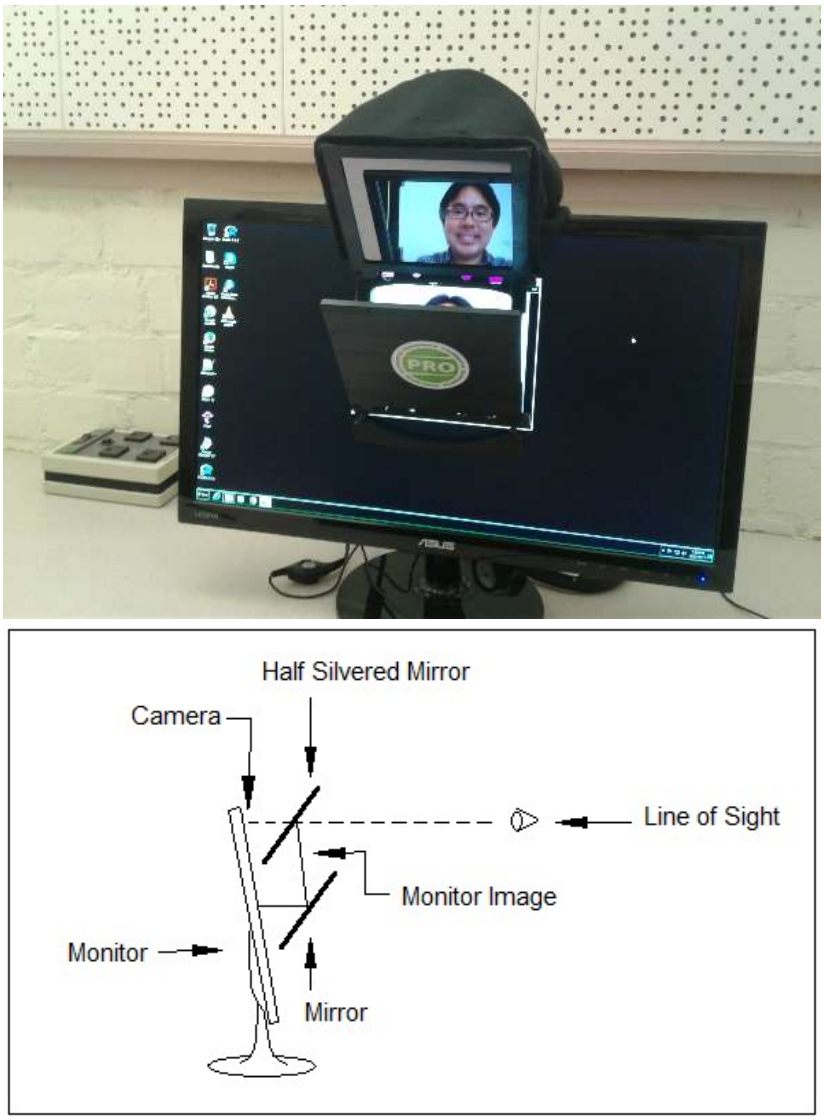


Figure 3.1. ProPrompter Desktop device in use for the current study (above). Basic concept design of how the apparatus corrects line-of-sight (below).

In sum, the main difference between the four training conditions was the type and number of FTF elements provided by each setup, namely auditory and/or visual information, synchronicity, and eye contact affordance. For the video condition, audiovisual information of the interlocutor (the tutor) was provided via video recording, while the audio-only condition provided only auditory information. The videoconferencing condition provided synchronicity in addition to the audiovisual

information of the video condition, and the videoconferencing with eye contact condition further added the ability of direct eye gaze.

(4) Posttest. After the completion of the second lesson, participants were given the posttest. The posttest was administered in the same way as the pretest, and contained the same tasks, instructions, practice trial, and test trial (test items were presented in a randomized order different from the pretest).

(5) Delayed Posttest. Approximately one week after the posttest, a second, delayed posttest was administered to the participants. This posttest was administered in the same way as the first posttest and pretest, and contained the same tasks, instructions, practice trial, and test trial (test items were presented in a randomized order different from the pretest and first posttest).

(6) Exit Survey. Once the responses for the delayed posttest were recorded, participants then completed the exit survey (see Materials, Survey 2). Upon completion of the exit survey, participants were debriefed and compensated with course credit.

3.3 Research Questions and Hypotheses

There were two research questions, followed here by the predictions for each:

1. What are the cognitive learning gains, as measured by error rates and reaction times, of various mediation types used for the acquisition of pitch accent? To what extent will these gains differ between audio-only (asynchronous), video (audiovisual asynchronous), and videoconferencing (audiovisual synchronous) conditions?

Some learning was expected in the audio-only condition, due to the training effects of the lessons and to initial accommodation effects, and this would be reflected in lower error rates in the posttests as compared to the pretest. The addition of visual

information in the video condition was expected to result in a lower posttest error rate as compared to the audio-only condition. The further addition of synchronicity in the videoconferencing condition was expected to result in a lower posttest error rate as compared to the video condition, as well as quicker reaction times. It was further expected that the more FTF-like the mediation was, the more retained learning would occur, as shown in the delayed posttest.

2. If the additional social cue of mutual eye gaze is added to a videoconferencing interaction, does this significantly increase learning gains due to higher perceived social presence?

It was predicted that the addition of direct eye gaze to the videoconferencing environment would lead to further learning gains as reflected in lower error rates and better retention when compared to the videoconferencing condition without possible direct eye gaze. This would serve as evidence that direct eye gaze allowed for the greater perception of social presence in this mediated environment. However, it may be possible that the converse is true and non-mutual eye gaze would decrease perceived social presence due to being distracting, as observed by Fullwood and Doherty-Sneddon (2006). If this occurred, it was expected that the error rates and retained learning of the eye gaze enabled condition would still show more gains than the condition without eye gaze, but that these results would be closer to those of the video and audio-only conditions. The results of the condition without eye gaze would also reflect this by being significantly worse (higher error rates and less retention) than those of the audio-only and video conditions.

There was also the possibility that some participants may have felt less comfortable knowing that they could be seen by the tutor, thus increasing their anxiety. If this was the case, then some participants in the videoconferencing conditions may have performed worse in their cognitive learning than those who could not be seen by the tutor in the prerecorded conditions. This would be reflected in higher error rates and reaction times that correspond with answers to the second part of the exit survey that indicate less comfort in being seen by the tutor.

3.4 Conclusion

This chapter gave an overview of the current study, including detailed descriptions of the materials used. The surveys gathered a varied but targeted selection of information, chosen for its potential to add further insight to the main collected data sets of error rates and reaction times from the three test sessions. The lessons were designed to reflect exercises that could be part of supplemental work for a language learning course. These were presented using four possible conditions that an instructor could set up for students, including a videoconferencing session with more realistic eye contact accomplished with a device that requires no additional software. The procedures for the current study were described in detail, and predictions for the research questions being addressed were also covered.

Chapter 4: Data Analysis and Results

4.1 Data Analysis

The primary data collected was the error rates and reaction times of each test item in both test tasks, for the pretest, posttest, and delayed posttest. This data formed the primary analysis for the study, and will be addressed first. The data collected in the two surveys was also analyzed for its possible clarification of the test task data results.

4.1.1 Test error rates and reaction times. For each condition, the mean error rates and reaction times were calculated. For ease of interpretation, the mean error was converted to mean accuracy, or mean frequency of correct answers. For the same/different discrimination task only, the baseline of this data (the pretest accuracy rate) was subtracted out from each participant's posttest and delayed posttest accuracy rate, resulting in the accuracy gains in pitch placement perception that each participant achieved as a result of completing the lessons.

The reaction time means were calculated differently, in that only the reaction times for correctly answered items were included. For the same/different discrimination task only, these correct-only means were then treated similarly to the accuracy rates, with the pretest mean reaction time (the baseline) subtracted out from each participant's posttest and delayed posttest reaction time mean, yielding the gains (or losses) in speed of correct pitch placement perception for each participant as a result of the lessons.

These data were then compared using repeated measures ANOVA tests (for each of the two test tasks) to explore main effects and interactions, with pitch placement perception as the within-subjects factor, and mediation type as the between-subjects factor. Planned pair-wise post-hoc comparisons using *t* tests were performed to further

explore differences between the four mediation conditions and between the test occurrences (time of testing). An alpha level of .05 was used for all statistical tests.

4.1.2 Survey data. Pearson's correlation coefficients were calculated for the items in the two surveys, both with the other items of the surveys as well as the accuracy gains and speed gains of the task 1 test data from the four conditions. The correlations of the survey items with other items from the same survey section were calculated to verify the validity of the two main exit survey sections (situational motivation and comfort with tutor presence) as well as to detect possible expected and unexpected relationships with other survey items. Finally, correlations for the survey items with the test data were calculated as a way to both clarify possible result patterns as well as detect potential confounds (such as frequency in use of videoconferencing before the experiment, for example). An alpha level of .05 was used for all correlations.

4.2 Results

4.2.1 Test accuracy and speed. The hypothesis was that while some learning was expected in all the mediation conditions due to the training from the lessons, greater accuracy gains would result in the conditions with more FTF-like features afforded; audio-only was expected to show the least gain and retention, and videoconferencing was expected to show the most. Between the two videoconferencing conditions, the one with direct eye contact was expected to show greater gains and retention due either to this social cue creating social presence or to the lack of it being distracting in the condition without eye contact. The participants' mean gains in accuracy and speed of pitch placement detection were analyzed using repeated measures ANOVA tests, starting with the data for the same/different discrimination task. For this first task, the mean accuracy

rates and reaction times at the three times of testing for each condition are shown in Tables 4.1 and 4.2.

Table 4.1

Mean Accuracy Rates for Task 1 at Times of Testing

Condition	Pretest	Posttest	Delayed posttest	*n*
Audio-only	.676	.804	.800	22
Video	.692	.765	.733	21
Videoconferencing	.732	.847	.893	19
Videoconferencing with eye contact	.690	.714	.759	20

Table 4.2

Mean Reaction Times for Task 1 at Times of Testing

Condition	Pretest	Posttest	Delayed posttest	*n*
Audio-only	763	785	669	22
Video	845	664	553	21
Videoconferencing	707	621	531	19
Videoconferencing with eye contact	890	825	734	20

Note: Rates for correct answers only; reported in milliseconds.

The data here shows no speed/accuracy trade-off pattern between the mean accuracy rates and response times, so further analysis focused primarily on differences and gains in accuracy between conditions at the times of testing. As described earlier, the error rates were converted to mean accuracy rates for ease of interpretation, and any gains in accuracy on the posttests were further calculated using these means. The resulting mean accuracy gains for this task can be seen in Figure 4.1.

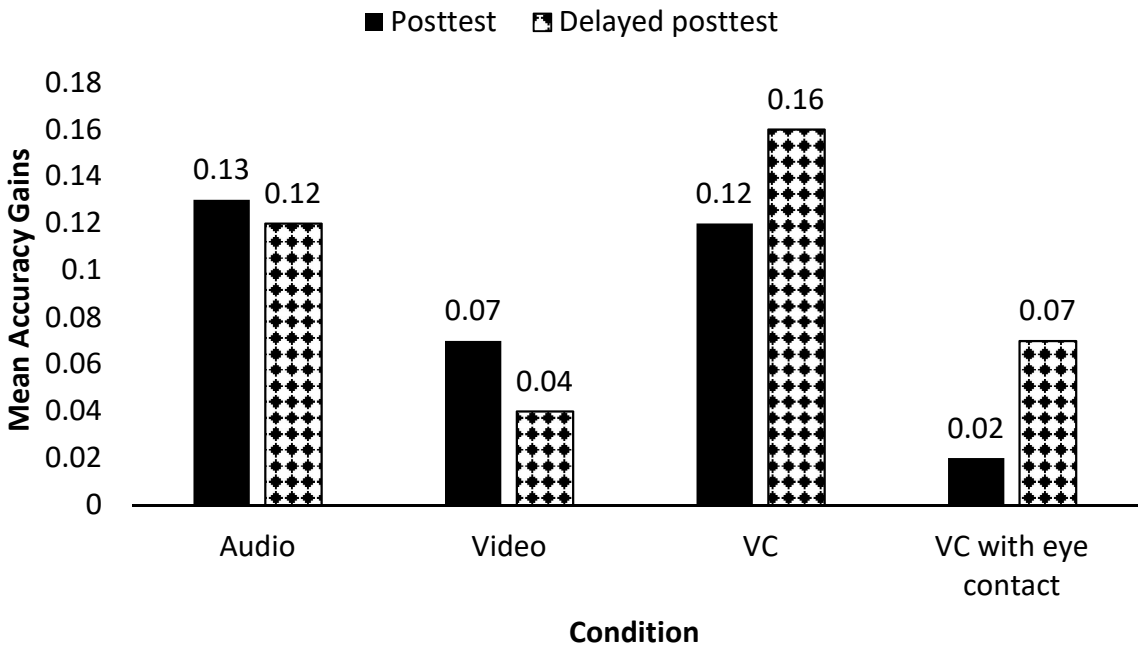


Figure 4.1. Mean accuracy gains (gained frequency of correct answers) for the posttest and delayed posttest of the same/different discrimination task for all four conditions. VC = videoconferencing.

The ANOVA for the mean accuracy rates in the same/different discrimination task showed a main effect for time of testing ($F(2, 156) = 21.51$; $p < .0001$) such that learning of pitch placement improved from the pretests to both posttests. This shows that the two lessons were effective in training for better perception of the pitch placement in Japanese words. There was no main effect for the different mediation types used for administering the lessons ($F(3, 78) = 1.98$; $p = .123$). Contrary to the hypothesis, this result suggests that while learning is taking place, the type of mediation used for the lessons did not influence the learning of pitch placement perception.

Analysis was also done of the accuracy gains (with the baseline of the pretest accuracy rates removed) at each posttest separately to further investigate any possible effects for the different mediation types. In this analysis, there was no main effect for

mediation type at the first posttest ($F(3, 78) = 1.56$; $p = .207$). Planned pair-wise tests between the gains of each mediation condition showed no significant differences.

The main effect for mediation type for the delayed posttest gains was tending towards significance, however ($F(3, 78) = 2.35$; $p = .078$). Planned pair-wise tests revealed this was due to the videoconferencing without eye contact condition having significantly higher gains than the video condition ($t(35) = 2.61$, $p = .013$), and nearly significantly higher gains than the videoconferencing with eye contact condition ($t(36) = 1.87$, $p = .069$). This suggests that while more data may be needed to detect a possible effect, a potential effect of mediation may be present.

Additionally, a two-way interaction between all three times of testing and the four mediation types for the mean accuracy rates also have a result tending towards significance ($F(6, 156) = 2.02$; $p = .066$). The planned pair-wise tests of the four mediation conditions' accuracy gains (with the baseline pretest accuracy rates removed; see Figure 4.1) revealed a significant difference between the two videoconferencing conditions across both posttests ($t(75) = -2.50$, $p = .014$), with the condition with eye contact having a significantly lower mean gain in posttesting than the condition without eye contact. While the audio-only condition had a significantly higher gain across posttests than the videoconferencing with eye contact condition ($t(122) = 1.97$, $p = .051$), the videoconferencing without eye contact condition had a significantly higher gain across posttests than the video condition ($t(106) = 2.29$, $p = .024$), and had a slightly higher gain (though not significant) than the audio-only condition. One of the videoconferencing conditions outperforming the others across posttesting does imply a potential advantage for mediation types that incorporate audiovisual synchronous cues.

This suggests that with more data, the influence of the lessons themselves on pitch placement perception may show a dependency on the mediation type used to administer them. The results also showed a trend in the delayed posttest gains beyond those of the posttest, where the delayed posttests of the two videoconferencing conditions appear to have greater gains than the posttests, which did not occur for the audio-only and video conditions (though not significantly; $F(3, 78) = 0.61$; $p = .608$). This further suggests that more data may be needed to detect an effect on longer-term retained learning due to mediation type, specifically related to the presence of synchronicity.

The ANOVA for the reaction times in the same task also showed a main effect for time of testing ($F(2, 156) = 20.29$; $p < .0001$), further confirming that detection improved from the pretests to the posttests overall, with the exception of a slight loss (or slower reaction times) for the posttest of the audio-only condition. The delayed posttest of this condition reflects the faster reaction times to correctly answered items observed in the rest of the posttests for the other conditions. The planned pair-wise tests (with results collapsed across mediation conditions) showed the delayed posttests had significantly faster reaction times than the posttests ($t(159) = 2.40$, $p = .017$), possibly suggesting retained learning.

The reaction times were used to calculate mean speed gains/losses as well. For changes in speed of accurate pitch placement detection, there was no main effect for the different mediation types ($F(3, 78) = 2.42$; $p = .072$). However, the planned pair-wise tests of the four conditions showed that significant differences occurred between the audio-only and video conditions ($t(82) = 3.90$, $p < .0001$), and the videoconferencing with eye contact and video conditions ($t(63) = 1.97$, $p = .053$), with the video condition

having faster reaction times on the posttests in both cases. No other significant differences between mediation types occurred, with the overall trend of the delayed posttest performances being faster than the posttests, which corresponds with the mean accuracy gains results suggesting that more data may be needed to observe any influence of the meditation types on the learning of pitch placement detection.

Following the same/different discrimination task was the picture recognition task. For this task, the pretest was removed completely from the analysis, as the nature of the task (determining if a word correctly identified a picture or not) in combination with the participants tested (no prior experience using the Japanese language) meant that any meaningful performance on this task could only occur after the participants had received the lessons, which exposed the participants to the words in conjunction with their meanings and representative pictures. The mean accuracy rates and reaction times at the two times of posttesting for each condition are shown in Tables 4.3 and 4.4.

Table 4.3

Mean Accuracy Rates for Task 2 on Posttests

Condition	Posttest	Delayed posttest	*n*
Audio-only	.453	.491	22
Video	.489	.492	21
Videoconferencing	.451	.470	19
Videoconferencing with eye contact	.441	.477	20

Table 4.4

Mean Reaction Times for Task 2 on Posttests

Condition	Posttest	Delayed posttest	*n*
Audio-only	1172	1034	22
Video	1034	888	21
Videoconferencing	991	843	19
Videoconferencing with eye contact	1209	966	20

Note: Rates for correct answers only; reported in milliseconds.

Mean error and mean speed were examined for this task's posttests, using repeated measures ANOVA tests and planned pair-wise tests. The ANOVA for mean error in the posttests of the picture recognition task showed no main effects for either time of testing or mediation type, with the frequency of correct answers occurring at chance levels. This indicates that while the two brief lessons and short time frame used for this study were effective enough to train participants to recognize the pitch placement distinction, they were not sufficient to teach the vocabulary words used. No significant interactions were found.

Gains and losses in reaction times were not calculated as there was no reliable baseline (the pretest results had been removed due to being unmeaningful), but similarly to the first task, only reaction times for correctly answered items were included. The ANOVA for speed of reaction times in this task did show a main effect for time of testing ($F(1, 78) = 29.91$; $p < .0001$). The planned pair-wise tests (with results collapsed across mediation conditions) showed the delayed posttests had significantly faster reaction times than the posttests ($t(159) = -3.90$, $p < .0001$). This corresponds with the similar main effect found for the same/different discrimination task, showing reaction times on correctly answered items became faster a week after the training. Also, there was a main

effect found for mediation type ($F(3, 78) = 3.27$; $p = .025$), which planned pair-wise tests across posttests revealed to be due to the videoconferencing condition being significantly faster than the audio-only condition ($t(75) = 2.80$, $p = .006$) and the videoconferencing with eye contact condition ($t(73) = 2.55$, $p = .013$). The video condition was also significantly faster than the audio-only condition ($t(83) = 2.50$, $p = .014$) and the videoconferencing with eye contact condition ($t(77) = 2.20$, $p = .031$). This is similar to the data trend from task one, in that the audio-only and videoconferencing with eye contact conditions showed the overall slowest reaction times. This also indicates that more data may be able to further clarify any influence on performance by the different mediation types. However, as the accuracy of answers for this task on both posttests were at chance levels, these reaction time results do not reflect meaningful information on the potential influence of the mediation type on the participants' learning, and will not be discussed further.

4.2.2 Survey results. Correlations of the demographic survey with test accuracy results were calculated to further explore any relationships or confounds between the test results and the participants' demographics, vision or hearing problems, musical training, exposure to/use of tonal and non-tonal languages, and familiarity with the mediation technologies used in the study. No significant correlations were found for any of the demographic survey items.

Correlations of the exit survey responses with test accuracy results were calculated to further explore the possible effects of situational motivation influences on learning within the different mediation conditions. Only the results of the first task (same/different discrimination task) were used, as only this task showed reliable accuracy

gains from the pretest to the posttests. Additionally, data from participants who showed an accuracy loss from the pretest to the posttest or delayed posttest was removed, so that any significant correlations found would inform specifically on accuracy gains (posttest: 27 total removed, delayed posttest: 17 total removed). While no main effect was found for mediation type as an influence on accuracy rates, the two-way interaction between time of testing and mediation type that is tending towards significance suggest that there may be some potential influence due the mediation type. The exit survey gathered data related to the participants' potential motivation types for completing the lessons, as well as their comfort in having a tutor able to watch them work in real time. Any correlation this data may have with accuracy gains occurring in the different mediation conditions could further clarify the test results.

To explore this, Cronbach's alpha internal consistency reliability coefficients were calculated to verify that the items were detecting the same subscales of situational motivation as originally designed for by Guay et al. (2000). There were high to moderate levels of internal consistency for each subscale: intrinsic motivation, $\alpha = .820$ (four items); identified regulation, $\alpha = .792$ (four items); external regulation, $\alpha = .774$ (four items); and amotivation, $\alpha = .684$ (four items). Given this, the motivation survey items were then correlated with the mean accuracy gains of participants who showed no accuracy loss in each of the mediation conditions.

When these correlations for each mediation type's posttest and delayed posttest were compared, items intended to detect intrinsic motivation (item 6) and identified regulation (items 3, 15) showed moderate to strong negative correlations with the gains in the videoconferencing with eye contact condition, and an identified regulation item (item

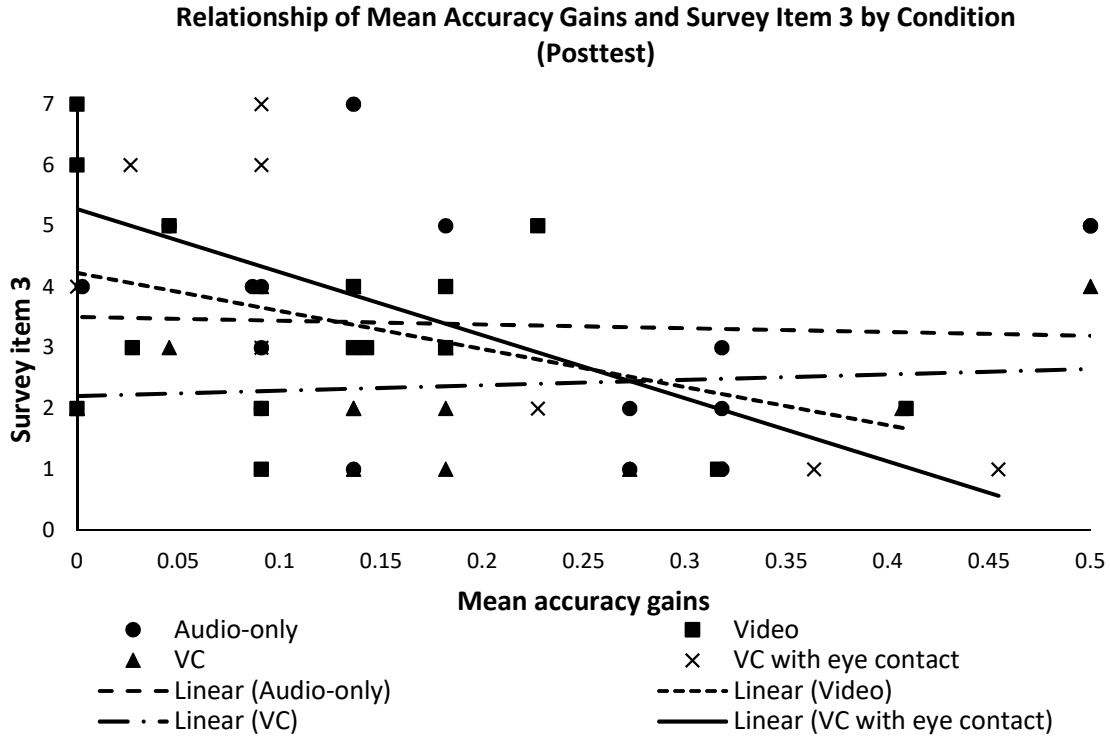
15; see Appendix A for survey items) showed moderate negative correlations with the video condition, while items for external regulation and amotivation did not correlate with the gains of any condition (see Table 4.5). This shows that within the videoconferencing with eye contact condition, the higher participants scored on these indicators of intrinsic motivation and identified regulation, the less gains they made in their accuracy of detecting pitch placement distinction. This is also shown for the video condition and an indicator of identified regulation (see Figures 4.2 and 4.3).

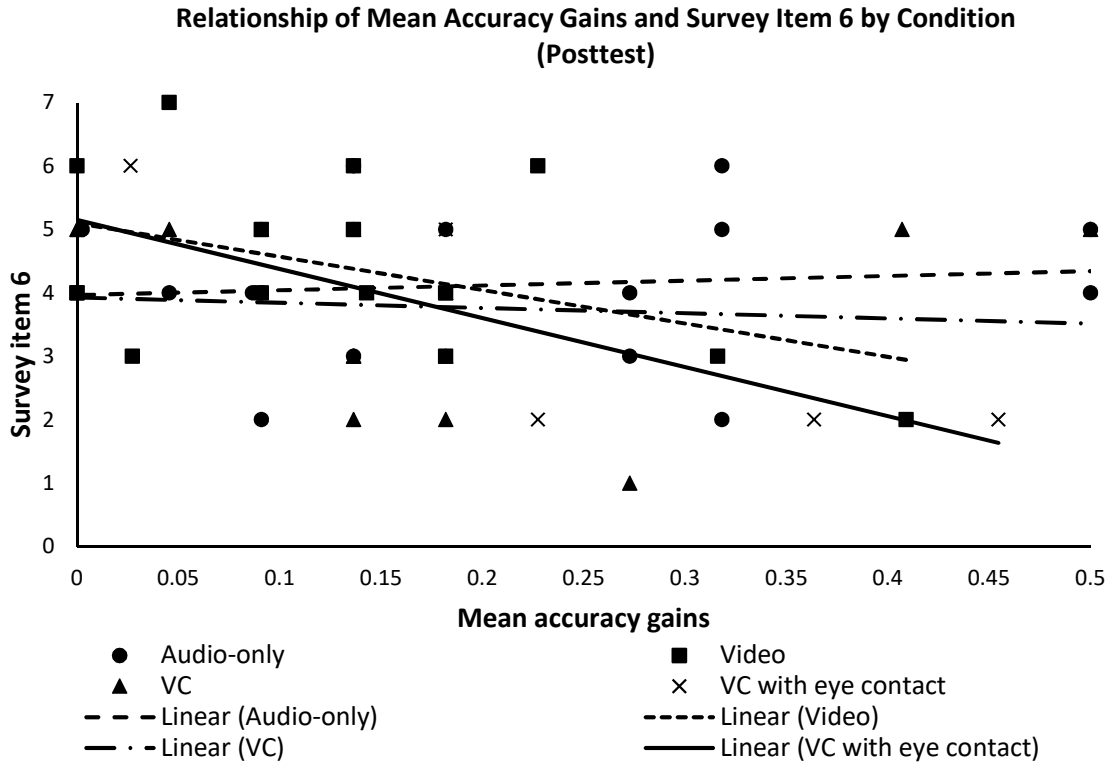
Table 4.5

Correlations of Relevant Survey Items to Mean Accuracy Gains of Videoconferencing with Eye Contact and Video Posttests

	Videoconferencing with eye contact		Video	
Survey items	Posttest	Delayed posttest	Posttest	Delayed posttest
3	-.70*	-.56*	-.42	-.43
6	-.79**	-.76**	-.45	-.44
15	-.70*	-.54*	-.53*	-.64**

*$p < .05$. **$p < .01$.


Relationship of Mean Accuracy Gains and Survey Item 3 by Condition (Posttest)
Survey item 3
Mean accuracy gains
Audio-only
VC
Linear (Audio-only)
Linear (VC)
Video
VC with eye contact
Linear (Video)
Linear (VC with eye contact)


Relationship of Mean Accuracy Gains and Survey Item 6 by Condition (Posttest)
Survey item 6
Mean accuracy gains
Audio-only
VC
Linear (Audio-only)
Linear (VC)
Video
VC with eye contact
Linear (Video)
Linear (VC with eye contact)

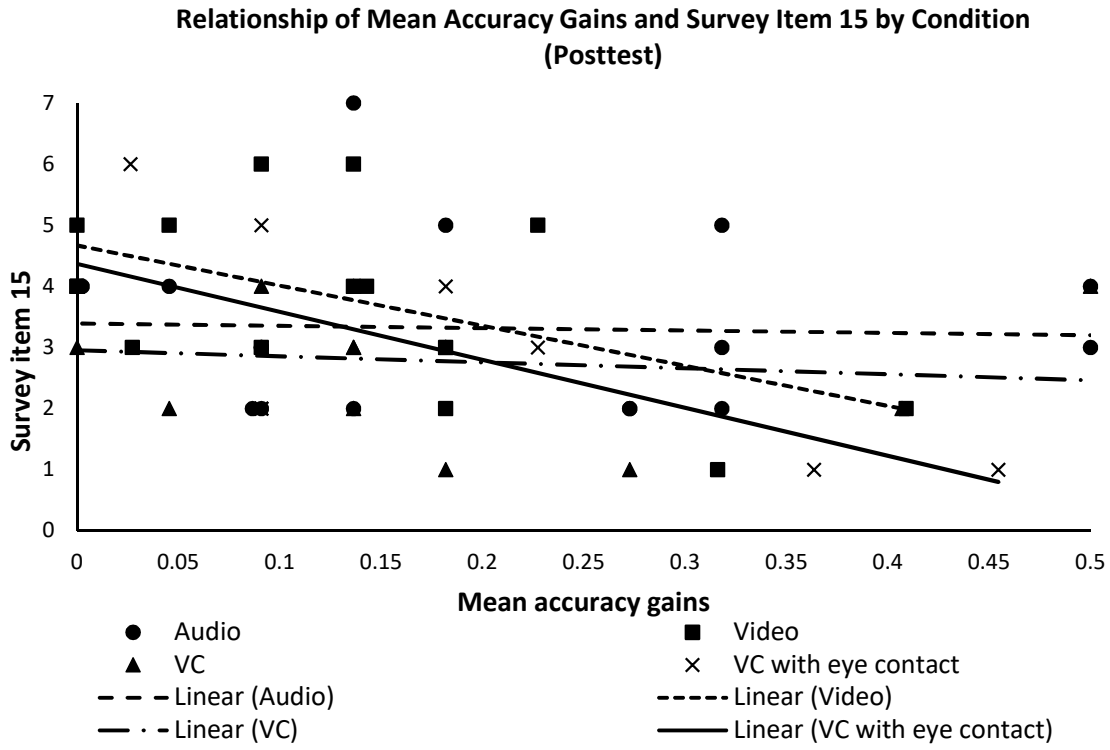
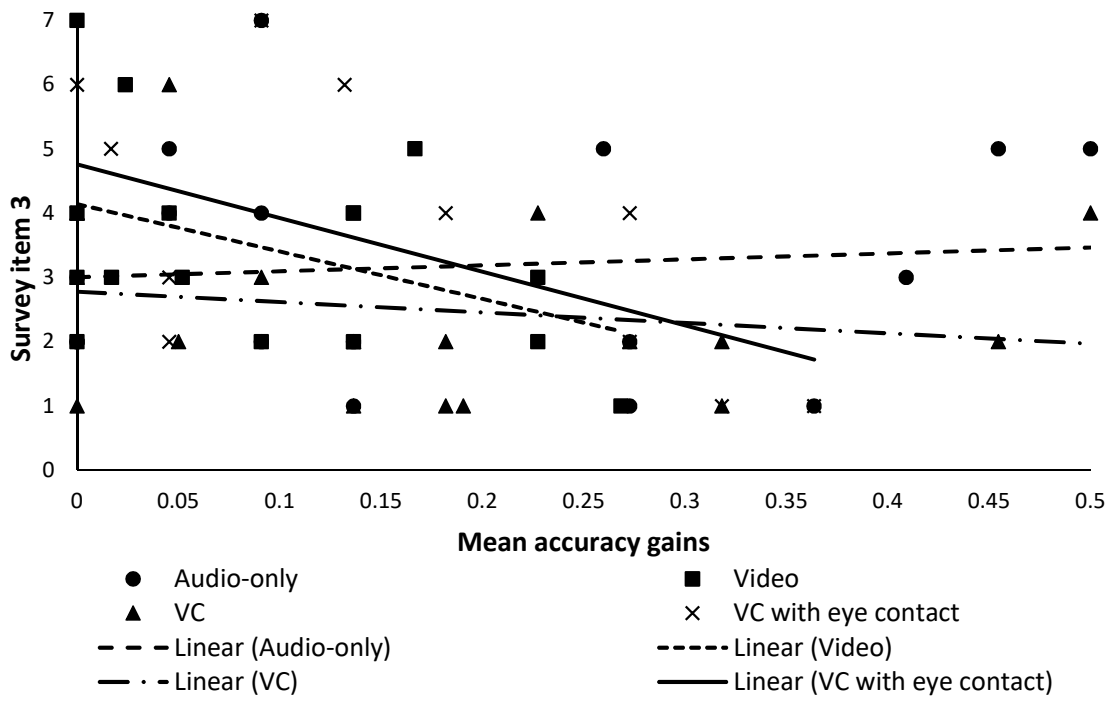
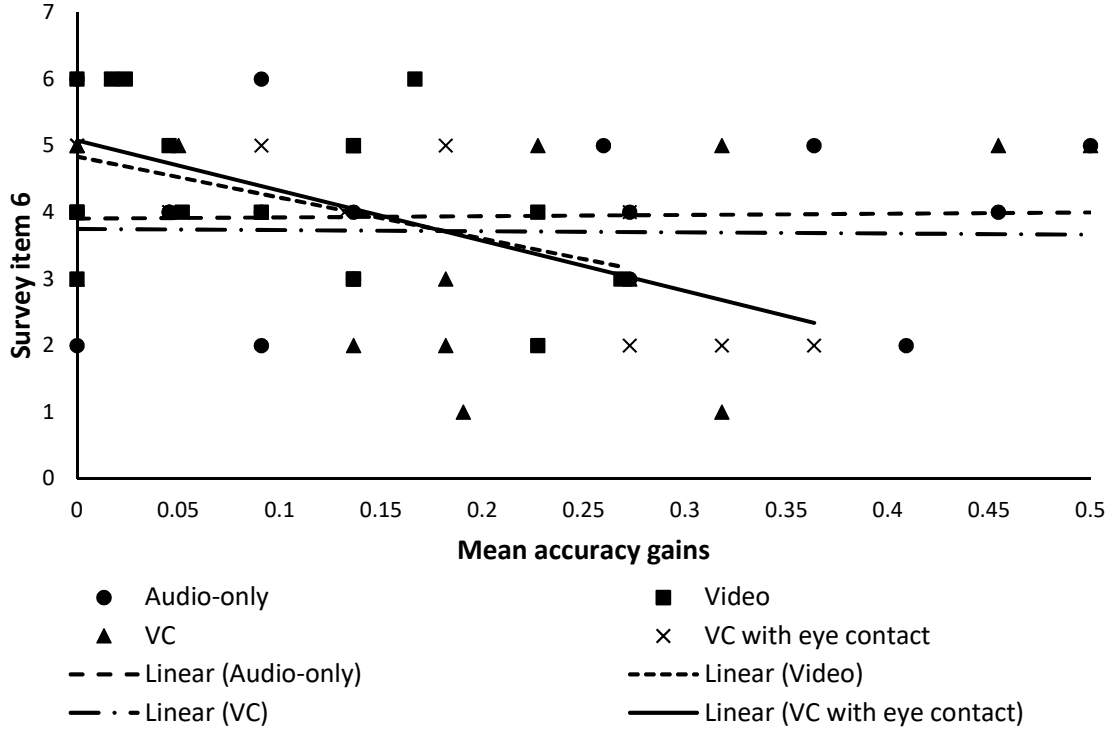


Figure 4.2. Scatterplots of the correlational relationships between the mean accuracy gains of each condition's posttest with motivation survey items 3, 6, and 15. Only participants showing no accuracy loss are included. VC = videoconferencing.

Relationship of Mean Accuracy Gains and Survey Item 3 by Condition (Delayed Posttest)



Relationship of Mean Accuracy Gains and Survey Item 6 by Condition (Delayed Posttest)



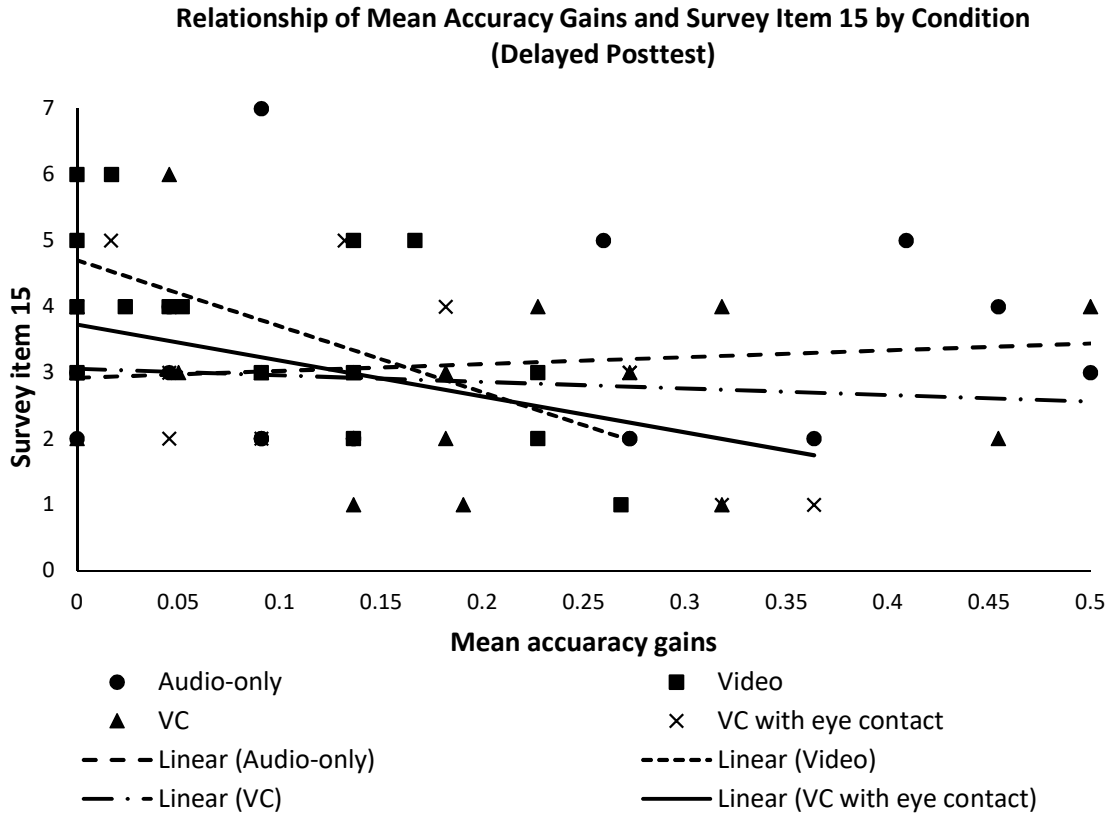


Figure 4.3. Scatterplots of the correlational relationships between the mean accuracy gains of each condition's delayed posttest with motivation survey items 3, 6, and 15. Only participants showing no accuracy loss are included. VC = videoconferencing.

The similar relationships of the intrinsic motivation and identified regulation items to the testing data suggested that the instrument may be detecting the participants' self-determination, as observed and discussed by Guay et al. (2000). Based on self-determination theory (Deci & Ryan, 1985), Guay et al. confirmed through multiple experiments that intrinsic motivation and identified regulation correlate with each other in this instrument as a reflection of higher self-determination, while external regulation and amotivation correlate with each other to reflect lower levels of self-determination. Guay et al. defined self-determination as "a true sense of choice, a sense of feeling free in doing what one has chosen to do" that is best expressed on a continuum from high to low levels of this sense (p. 176). The results here did show items for both intrinsic motivation

and identified regulation (an extrinsic motivation measure) correlating similarly to the same conditions. With these results suggesting that the survey was detecting a possible self-determination vs. lack of self-determination relationship with the mediation condition test results, the survey item responses were collapsed into two main subscales. Intrinsic motivation and identified regulation items were grouped into a "high self-determination" mean score, and external regulation and amotivation were grouped into a "low self-determination" mean score for each participant. These two new subscales showed high internal consistency, as determined by a Cronbach's alpha of .878 for high self-determination (eight items) and .812 for low self-determination (eight items).

These two subscales were then correlated with the accuracy gains of the four mediation conditions. This was done separately for the posttest and delayed posttest results, so that any differences between the two times of testing after the training could be observed. High self-determination and videoconferencing with eye contact were strongly negatively correlated for both the posttest ($r(8) = -.80$, $p = .005$) and the delayed posttest ($r(12) = -.69$, $p = .006$). Following the pattern found with the survey items before they were collapsed into these subscales, when participants had greater self-determination (or feeling that they had freely chosen to do these lessons), the less accurate they were at detecting the pitch placement distinction (see Figure 4.4). High self-determination did not significantly correlate with any other mediation condition, for either posttest. Low self-determination also did not show significant correlation with any mediation condition for either posttest.

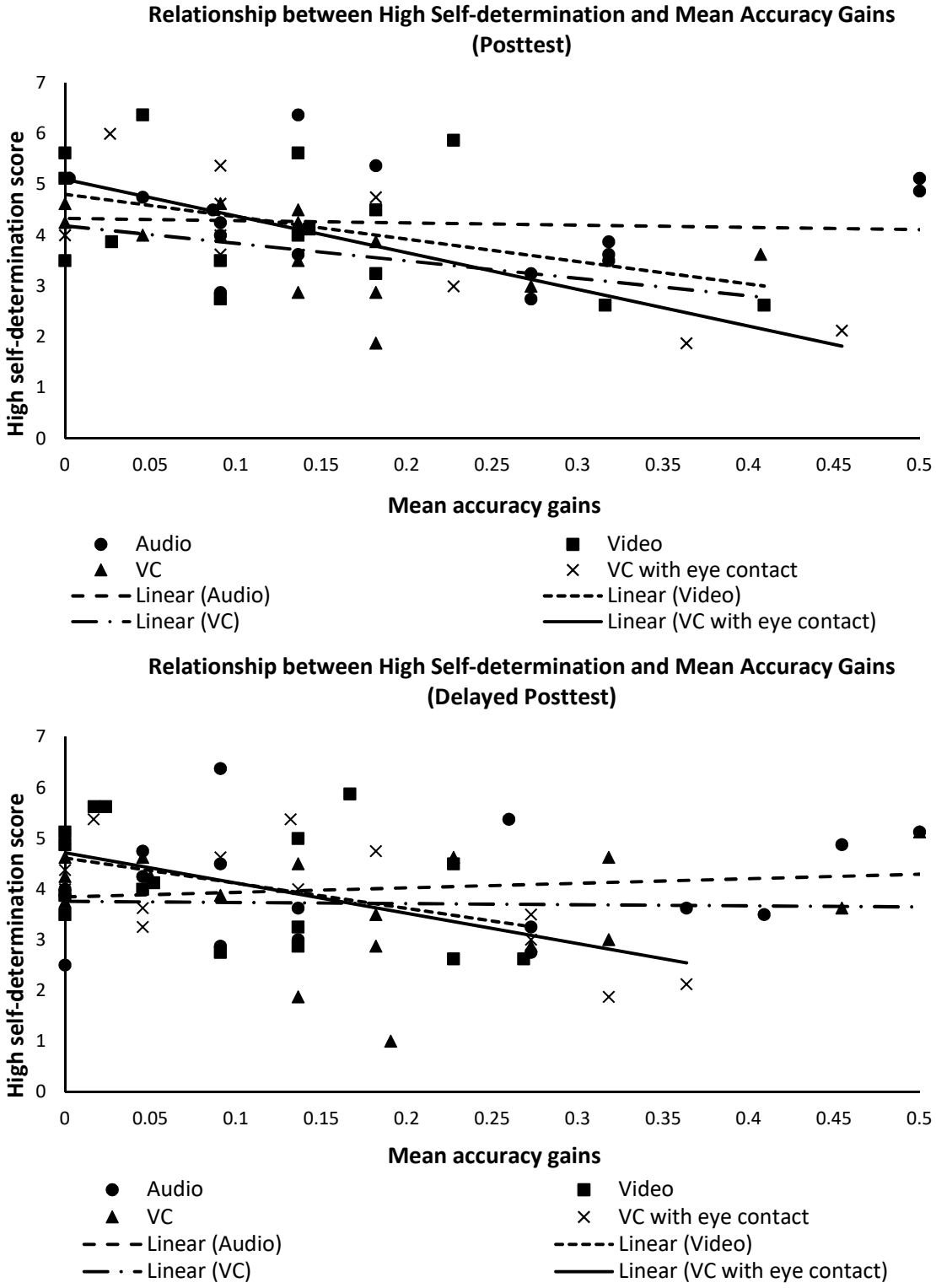


Figure 4.4. Scatterplots of the correlational relationships between the high self-determination score and the mean accuracy gains for each condition's posttest and delayed posttest. Only participants showing no accuracy loss are included. VC = videoconferencing.

The tutor comfort survey items were collapsed into two groups: "tutor presence desirable" and "tutor presence not desirable." While these two groups had high internal consistency, with a Cronbach's alpha of .789 for the former (four items) and .765 for the latter (three items), there was no discernable correlational pattern of either group with a particular mediation condition or with the self-determination groups.

The final two survey items in the exit survey asked about participant's plans for possible future study of the Japanese language. The first item asked about likelihood of enrolling in a course to learn Japanese (item 1), and the second asked about likelihood of using self-study to learn Japanese (item 2). This scale had a moderate level of internal consistency, as determined by a Cronbach's alpha of .725. The answers for these items in each mediation condition were correlated with the self-determination subscales from the situational motivation survey, using the data from participants who experienced accuracy gains on the delayed posttest, which was administered immediately before the exit survey. Overall, the likelihood of either method of study tended to correlate positively with high self-determination, and negatively or not at all with low self-determination (see Table 4.6). This indicates that for participants in all of the conditions, the greater their sense of self-determination towards these lessons, the greater their likelihood of continuing to freely chose to study the language will be, especially through self-study methods. The inverse of this was also confirmed by the audio-only group; the lower their self-determination, the less likely they are to choose to continue further self-study of Japanese (see Figures 4.5 and 4.6).

Table 4.6

Correlations of Japanese Study Survey Items to Self-Determination Scores by Condition

Condition	Survey item	High self-determination	Low self-determination
Audio-only			
	1	-.03	.06
	2	.54*	-.55*
Video			
	1	.53*	-.40
	2	.49*	-.33
Videoconferencing			
	1	.26	.10
	2	.51*	-.18
Videoconferencing with eye contact			
	1	.55*	-.14
	2	.61*	-.32

*$p < .05$.

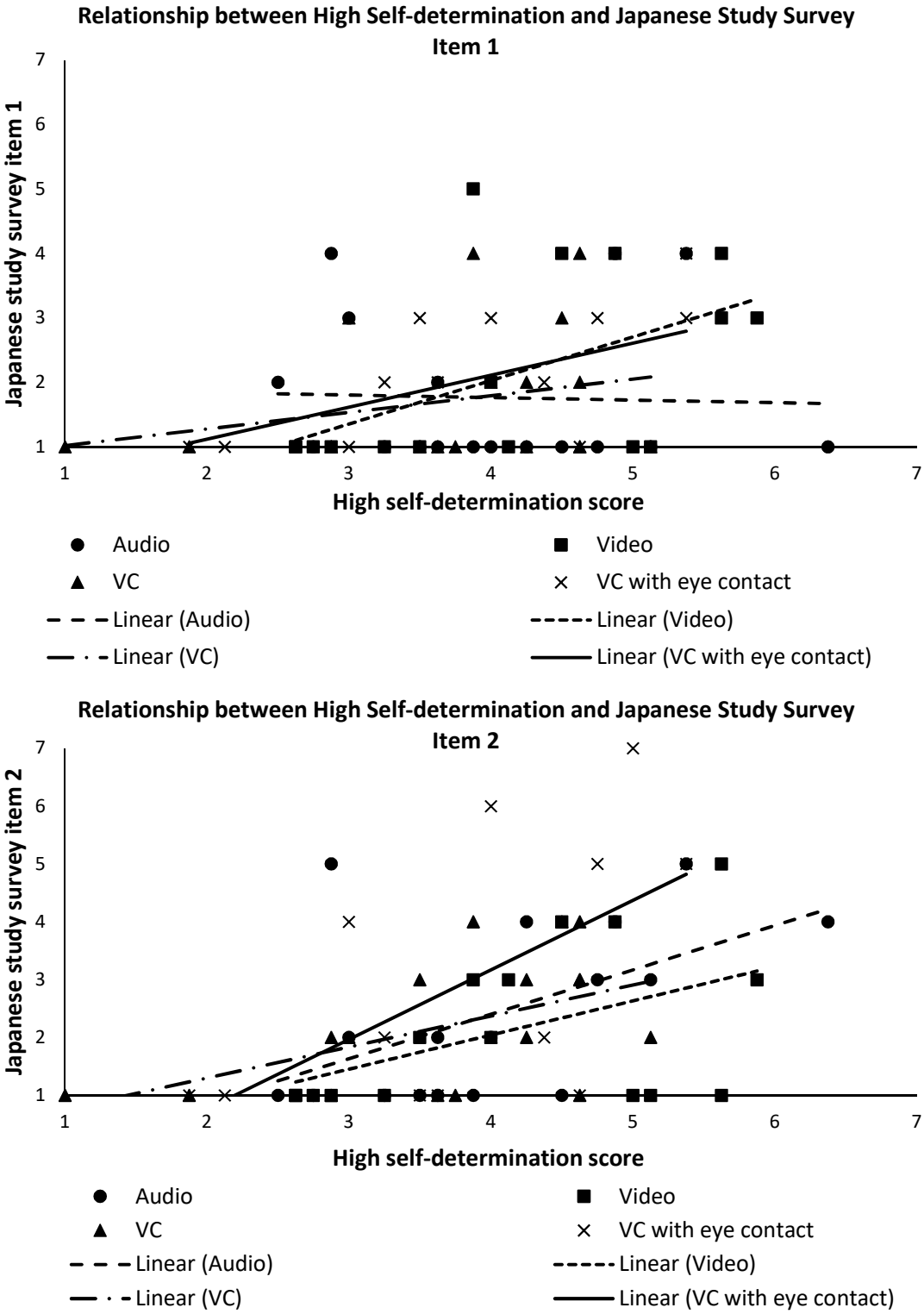


Figure 4.5. Scatterplots of the correlational relationships between the high self-determination score and the responses to Japanese study survey items 1 and 2 for each condition's delayed posttest. Only participants showing no accuracy loss are included. VC = videoconferencing.

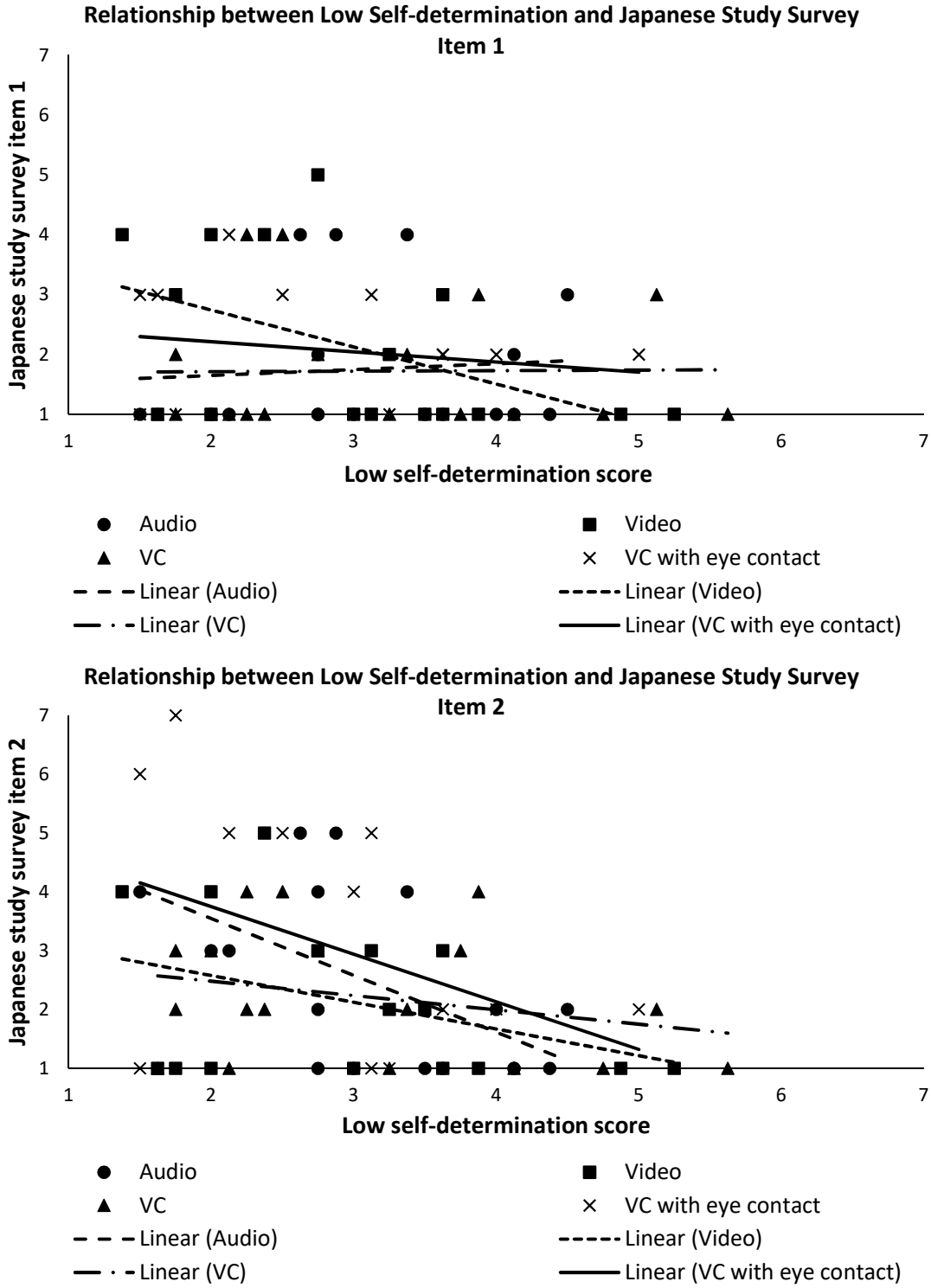


Figure 4.6. Scatterplots of the correlational relationships between the low self-determination score and the responses to Japanese study survey items 1 and 2 for each condition's delayed posttest. Only participants showing no accuracy loss are included. VC = videoconferencing.

4.3 Conclusion

This chapter covered both the analysis of the data and the results of that analysis. The main data set for the study was the error rates and reaction times for the pretest, posttest, and delayed posttest for the two testing tasks given to all four conditions. The error rates were converted to accuracy rates so that mean accuracy gains between tests could be compared. Reaction times were used to calculate mean gains or losses in response speed for correctly answered items. The data from the two surveys was analyzed using Pearson's correlation coefficients. When correlated with the accuracy gains, only participants who showed no accuracy loss were included.

The accuracy rates results showed a significant main effect for time of testing but no significant effect for mediation type in the same/different discrimination task, and no significant effects for the picture recognition task were found. While the lessons were effective for teaching pitch placement distinction, the prediction for differences in performance due to mediation type was not confirmed. Also, the lessons were not sufficient for teaching the vocabulary used in the tasks. The change in speed results showed a main effect for time of testing, which supports that the lessons were effective for teaching the distinction. No meaningful significant effect was found for mediation type within the reaction time data; no evidence of a speed/accuracy trade-off effect was found.

For the possible influence of eye gaze between the two videoconferencing conditions, the same/different discrimination task showed a two-way interaction between time of testing and mediation type that was tending towards significance. The planned pair-wise tests showed this was due to the videoconferencing with eye contact condition

having significantly less accuracy gains than the videoconferencing without eye contact condition, as well as the audio-only condition. The prediction that eye gaze would improve the gains in a videoconferencing setup was not confirmed.

The results for the survey data showed that no items from the demographic survey correlated with the accuracy gains or speed gain/loss data. The exit survey data was compared with the results of the same/different discrimination task only. The results of the situational motivation section of the exit survey showed that a self-determination subscale was being detected, so these survey items were collapsed into high and low self-determination subscales, which showed high internal consistency. When these subscales were correlated with the accuracy gains of the four conditions, it revealed that within the videoconferencing with eye contact condition the more self-determination a participant had, the less accurate they were at detecting the pitch placement distinction. This could indicate the influence of another factor that shares a negative relationship with self-determination and motivation, such as anxiety, which could cause worse testing performance; this will be discussed further in Chapter 5.

Lastly, the tutor comfort items on the exit survey showed no correlational pattern with the mediation conditions or with the self-determination subscales. The two items addressing future study plans of Japanese language showed that in all conditions the more self-determination a participant had, the greater their likelihood of continuing to self-study Japanese.

Chapter 5: Discussion and Conclusions

5.1 Introduction

The purpose of this dissertation was to investigate different mediation types for their abilities to facilitate L2 learning. The central question of to what extent will cognitive learning gains differ between the use of three potential mediation types was addressed by using a focused case study on gains in listening perception of the Japanese language pitch placement feature by English L1 speakers. This question was approached using measurable learning gains through error rates and reaction times in a pretest/posttest/delayed posttest design, as well as surveys to provide additional information. The mediation types used were selected for their ability to facilitate various contextualization cues that could potentially foster perceptions of social presence. The presence or absence of direct eye contact as a further facilitating social cue was also investigated. Below is a review of the findings, such as the robustness of accommodation and social presence effects through the use of technology, as well as the influence of a student's own motivations for L2 learning. Limitations of the study, implications for second language learning, and suggestions for future study will also be discussed.

5.2 Potential Influence of Mediation Type on Speech Perception

The first research question of the study was: *What are the cognitive learning gains, as measured by error rates and reaction times, of various mediation types used for the acquisition of pitch accent? To what extent will these gains differ between audio-only (asynchronous), video (audiovisual asynchronous), and videoconferencing (audiovisual synchronous) conditions?* The hypothesis was that the closer a mediation type was to a FTF encounter, the lower the error rates and the faster the reaction times would be in the

posttests. While this was not confirmed, the results do suggest that all of the mediation types used could generate a degree of social presence and accommodation for the participants. For the same/different discrimination task, participants in all conditions experienced gains in pitch placement perception, showing that the brief lessons were effective. While there was no main effect of mediation type, the planned pairwise comparison results did show an effect suggesting that videoconferencing without eye contact may lead to more gains in pitch placement perception, and all four types did successfully facilitate some learning gains. This indicates that for learning to listen for a feature as subtle as pitch placement, short term training can be effective. The participants came from a linguistic background that places no emphasis on pitch as a feature that may determine difference in meaning, but even through the least FTF-like mediation type, they were able to learn to listen for this distinction as a minimal pair marker. This shows that technology can serve as a successful supplemental tool for teaching awareness of such a feature when access to native or near-native speakers is unavailable otherwise. This confirmation of the usefulness of technology in a particular case of adult L2 learning is encouraging, but potential differences between the different technologies used is still an area of interest.

While the picture recognition task did not show any effects for either time of testing or mediation type, this is not unusual considering the difference in the nature of the two tasks. The picture recognition task had the additional requirement of learning to correctly identify the vocabulary words with their meanings (represented pictorially), beyond also listening for the pitch placement distinction. It appears that the limited time frame and exposures of the study were not sufficient for this level of learning, which

requires forming more distinct mental representations of individual L2 words and is arguably more cognitively taxing than learning a new sound distinction rule. A longer study with more lessons designed for word learning and more exposures to the tutor will be needed to further investigate and determine if a difference between mediation types due to social presence impacts this task.

One of the factors important in determining the potential usefulness of a particular technology for language learning is its ability to create the perception of social presence in learners. As discussed in Chapter 2, social presence is impacted in both the nature and purpose of an interaction through various types of technology or mediation (Short et al., 1976). Social presence is assumed to be important when learning language, which is largely a medium for social behavior (Guichon & Cohen, 2014; Zhan & Mei, 2013). The results of this study suggest that further study may reveal different impacts on that perception of social presence. The results show that time of testing was significantly different between the pretest and posttests, with both error rates and overall reaction times decreasing in the same/different discrimination task. The training was effective for improving pitch placement perception through all four mediation types. Although no main effect was found for error rate improvement between the four mediation types, an interaction between time of testing and mediation type was tending towards significance, with some significant differences shown in the planned pair wise tests between the conditions. Specifically, the videoconferencing without eye contact condition was shown to improve the most by the planned pair wise tests, outperforming the video and videoconferencing with eye contact conditions. The audio-only condition also improved performance more than the videoconferencing with eye contact condition. No significant

difference was found for reaction times between the four mediation types as well, and the pattern of the data did indicate that no speed/accuracy trade-off effect was impacting the error rate results.

While solid conclusions of potential differences in effectiveness cannot be made from these results, they indicate that differences may be observable if more time and exposures to the tutor are afforded. Further research using a longer time frame with more exposures to training with the tutor could show if the possible differences suggested by the pairwise tests are evidence of stable effects. The overall trend did indicate a possible advantage in using synchronous videoconferencing over video and audio-only exposure, as that condition did result in a higher accuracy gain over the other two in the delayed posttest. While audio-only did have higher accuracy gains over the video condition, which was not expected, the design of the training and test materials may be the reason. As the test materials were administered in a way very similar to the audio-only condition (prerecorded audio clips with no visual information), this may have been somewhat advantageous in that the training reflected the testing more closely than the other conditions. Further, there is the possibility that audio-only was less distracting, in that it could have been easier for participants to focus on the sounds they were hearing.

Another possible reason for this result pattern may be that visual information did not have as strong an impact for this particular sound feature. While visual speech information such as differing lip movements can be useful for listeners when distinguishing between certain phonemes (Navarra & Soto-Faraco, 2007), any visual difference between the different pitch placements may have been too subtle to aid detection. When compared only visually (with the sound muted), the video of the tutor in

the current study pronouncing two members of a minimal pair showed no easily or obviously discernable difference between the words. Because the distinction was based solely on pitch placement, visual information may not have played a role to the extent that it can for other sounds features which form minimal pairs in language.

If synchronous videoconferencing provides the possible advantage of greater social presence, audio-only provides the possible advantage of easier focus or similarity to the testing materials, and visual information alone does not provide further advantage in pitch placement detection, the overall pattern seen here is reflective of these various impacts. It is further suggested that the addition of eye contact may be disadvantageous, as the videoconferencing with eye contact condition had lower accuracy gains than the other three conditions.

5.3 Potential Influence of Eye Contact on Speech Perception through Mediation

The second research question of the study was: *If the additional social cue of mutual eye gaze is added to a videoconferencing interaction, does this significantly increase learning gains due to higher perceived social presence?* The hypothesis was that the videoconferencing condition with eye contact would have lower error rates and reaction times than the videoconferencing condition using the standard setup that does not facilitate eye contact, but this was not confirmed. Furthermore, the more surprising suggestion of these results is that the videoconferencing with eye contact condition may lead to worse performance than the other mediation types, despite being the most FTF-like given its amount of available social information. This could indicate possible differences in the social presence generated by the conditions, because of the potential impact of anxiety.

Foreign Language Classroom Anxiety, more recently incorporated in the more comprehensive term of Foreign Language Speaking Anxiety (FLSA), was first defined by Horwitz, Horwitz, and Cope (1986) as a cluster of beliefs and perceptions held by language learners which show ties to communication apprehension, testing, and fear of negative evaluation anxieties. Specifically, it was noted that for L2 learning adults the experience of communicating in a language in which they are acutely aware of having limited ability can produce a certainty that it will be difficult, and these thoughts lead to anxiety. Adults think of themselves as competent in communication and sociocultural knowledge, but Horwitz et al. highlight that L2 speaking and listening in particular make this less certain, and "…any performance in the L2 is likely to challenge an individual's self-concept as a competent communicator and lead to reticence, self-consciousness, fear, or even panic" (p. 128). This can be further emphasized when an L2 learner is speaking to a native speaker of the L2, who is often considered a competent communicator by virtue of their fluency in the language. The awareness of a noticeable difference in fluency levels between these two speakers may aggravate FLSA, especially the fear of being negatively evaluated by the interlocutor.

Çağatay (2015) specifically investigated FLSA of L2 learners when speaking with native speakers of the L2. Through the use of a questionnaire, FLSA was measured for EFL students at four different proficiency levels in a Turkish university. While proficiency in the L2 was not found to affect FLSA levels, communicating with a native speaker in the L2 created significantly higher FLSA than speaking in front of fellow L2 learning classmates in the L2. Çağatay suggested that the students' often low opportunities for encountering native speakers socially may contribute to this increase, as

they may have an idealized image of native speakers' language use and feel a native speaker would be more critical than a fellow L2 learner.

FLSA is connected to social presence through its foundation of communication apprehension and fear of negative evaluation anxieties. These anxieties are defined by their social aspect – the ability to smoothly communicate and appear competent to other individuals. The perception of social presence is required for a failure of communication to be felt as a socially judgable error. Pertaub, Slater, and Barker (2002) found that even speaking in the much more familiar L1 (as compared to an L2) in clearly virtual environments, where the interlocutors are computer-generated animations, can be enough to trigger these anxieties.

Pertaub et al. (2002) investigated fear of social performance, specifically due to fear of negative evaluation. Participants were asked to prepare a brief (5 minute) talk on any subject of their choice, and then deliver this talk on two separate occasions to a virtual reality audience, composed of several animated men sitting around a conference table. Participants wore a head mounted VR device to see the animated environment, which was purposely animated to be representative but not highly realistic (in other words, the environment and the characters were clearly not real). Three possible "audience reactions" were randomly assigned to the participants: positive, where the characters sat attentively facing the participant while smiling and nodding in agreement; negative, where the characters had various actions of sleeping, slouching, leaving the virtual room, and facing away from the participant; and neutral, where the characters sat statically looking straight ahead. Through the use of several anxiety scales administered both before and after these talks, it was found that the emotional responses of the

participants to these virtual audiences matched the audience reaction they received, especially the negative scenario, where fear of negative evaluation anxiety was clearly triggered (Pertaub, Slater, & Barker, 2001; Pertaub et al., 2002).

The study by Pertaub et al. (2001, 2002) shows that social presence generated through mediated environments can be sufficient to trigger fear of negative evaluation anxiety, which is a major component of FLSA. In mediated environments such as the ones used in the current study, where the interlocutor was a real individual and a native speaker of the L2, it is reasonable to infer that FLSA may have had an influence on the participants' performances. Furthermore, Pertaub et al. (2001, 2002) assert that the purposeful use of simulated eye contact in the positive and negative scenarios of their virtual audience was important to establishing the social presence generated. The animated characters in both of these scenarios would make or break eye contact by looking at or away from the screen (the location of the participants' eyes in the VR headset). Also, the characters' eyes were designed so that their gaze direction was easy to detect.

The hypothesis of the current study was that the presence of eye contact, through its function as a contextualization cue in FTF communication, would increase the social presence generated by the videoconferencing condition and that this increase would be reflected in lower error rates and reaction times. However, it is possible that another effect is taking place – the increase of social presence is leading to an increase of FLSA. Increased FLSA would likely show as worse testing performance, as test anxiety is also a contributing component of the cluster of beliefs held by language learners that form their FLSA (Horwitz et al., 1986). In addition to this, the videoconferencing without eye

contact condition showed the highest gains, which may be due in part to the participants not expecting eye contact from the standard videoconferencing setup. This suggests that any social presence which was generated in an expected way (given the technology in use) was enough to be helpful, but not so overwhelming as to trigger anxiety. The fact that the videoconferencing with eye contact condition showed the least accuracy gains among all four conditions may be a reflection of its greater social presence unexpectedly leading to greater FLSA, and therefore worse performance results. The current study did not measure possible anxiety, but did find evidence of the phenomenon of self-determination being expressed in a way that may indicate the presence of FLSA.

The results of the adapted motivational survey items from the exit survey detected the participants' self-determination levels. Items from both the intrinsic motivation and identified regulation subscales showed similar correlational patterns, indicating that these items were functioning as a high self-determination subscale that was observed during the design and testing of the original instrument (Guay et al., 2000). Due to this, the survey items were collapsed into high and low self-determination subscales, which showed high internal consistency (see Chapter 4). When the means of these subscales were correlated with the posttest and delayed posttest results of the participants who showed learning gains, strong negative correlations were found between high self-determination and both posttests for the videoconferencing with eye contact condition. In other words, as self-determination increased for these participants, accuracy in detecting the pitch placement distinction decreased.

These results were unexpected, as they indicate that the higher a participant's self-determination was, the worse they performed in testing, specifically within the

videoconferencing with eye contact condition. Higher self-determination is associated with higher levels of intrinsic motivation due to the feeling of doing a task one has themselves freely chosen to do (Deci & Ryan, 1985, 2000). Higher intrinsic motivation to learn something is related to the learner placing a higher value upon what they are learning, and there is some evidence of a positive relationship between intrinsic motivation and learning outcomes for students using mediated environments (Cho & Heron, 2015).

The negative relationship found in the current study between self-determination and the accuracy gains of the most FTF-like condition at first seems counterintuitive, but could suggest that FLSA played a role. If a learner is more self-determined to learn a particular subject or task, they are placing more internal value on it and have a desire to perform well, especially when interacting with individuals who are considered more competent. Therefore, these results could reflect that the participants who had more self-determination towards the task and its subject matter would be more susceptible to the anxiety of not performing well in the opinion of a native speaker, and the greater social presence generated by the videoconferencing with eye contact condition made this anxiety all the more acute. These survey results further suggest that FLSA may be the reason that participants in this condition performed the worst for accuracy gains due to its ability to generate greater social presence.

Another explanation for the low accuracy gains in this condition could also be its novelty to the participants. Because the standard videoconferencing setup available to most people does not facilitate natural eye contact, many users have accepted this lack of eye contact as a normal characteristic of a videoconference. While eye contact that is

achieved by simply looking at the interlocutor's face is natural in a physically co-present FTF interaction, it may actually feel unusual in a mediated interaction, due to its current rarity. This novelty could have been distracting for the participants, thus leading to lower accuracy gains. Further study into this specific potential novelty effect is needed to confirm if this is indeed a possible reason for the results in the current study.

5.4 Potential Accommodation Effects of Speech Perception through Mediation

The evidence here which suggests that phonetic learning occurred in the four mediation conditions, combined with the greater effect in the videoconferencing condition, also suggests that social presence has an effect. Given the evidence of phonetic accommodation's automaticity and further promotion by social presence, it was predicted that the conditions which generate more social presence would also reflect more accommodation through greater accuracy gains.

As discussed in Chapter 2, some phonetic accommodation happens automatically in one's language use, including both L1 and L2 use, and occurs even in the least socially present environments, such as hearing single words as audio-only presentations in a lab (e.g., Babel & Bulatov, 2011). Often for L2 learners there is also the deliberate approach to accommodate their speech as reflected in attempts to produce native-like pronunciation of the L2. Flege (2007) has argued that an L2 learner's perception of L2 phonological segments impacts their ability to then produce the same segments accurately. In his Speech Learning Model (SLM), Flege explains that production of phonological elements of the L2 is guided by the perceptual representations stored in the learner's long-term memory, so an L2 learner must first perceive an element to produce it reliably and consistently in a generalized manner. Accommodation aids this process, as interlocutors

adjust their speech features to more closely match that of what they are perceiving in the other's speech, both phonologically and prosodically (Trofimovich, 2016). Accommodation effects provide evidence of perception of phonological segments, which is a crucial part of phonological learning.

Accommodation effects can be directly measured through speech production, but can also be indirectly measured through listening accuracy. Babel and Bulatov (2011) used a word repetition task to investigate participants' L1 accommodation to the fundamental frequency (F0) of tokens produced by a model speaker. Analysis of the degree of accommodation was done both through acoustic analysis of the participants' spoken repetitions, as well as through listening judgments by different participants who judged if the pre- or post-productions of the speaking group were more similar to the model tokens. The authors found that both methods of measuring accommodation were accurate and produced the same results; additionally, the listening judgment was viewed as a more holistic measure, since this allowed listeners to use any acoustic difference to make a judgment. This better represents natural speech, or the environment in which most judgements of speech quality are made by interlocutors. Because of this, the fact that the judgements of accommodation to a single feature were similar through both methods lead the authors to conclude that L1 accommodation is detectable by listeners (Babel & Bulatov, 2011).

Listener detection of F0 accommodation has also been found for L2 language use. Wang (2001) investigated English L1 and Norwegian L1 speakers' ability to accommodate to more native-like patterns of Mandarin tones, which differ in F0 values and fluctuations for each of the four tones. In two experiments (one using English L1

Mandarin L2 learners, the other using Norwegian L1 Mandarin L2 learners), a treatment group and control group were recorded speaking Mandarin words in all four tones both before and after a training period of 2 weeks. The treatment group received eight perceptual identification task sessions for the four tones. Both the English L1 and Norwegian L1 treatment groups improved in their ability to correctly identify the tones in the posttest, as well as two generalization tests using stimuli and voices not used in the training. Native Mandarin speakers judged the pre- and post-productions of the English L1 Mandarin L2 learners; the post-productions were rated significantly more accurate for tone production. The listener judgements were further confirmed with acoustic analysis of the pre- and post-productions, where the F0 values and contours of native speaker production were significantly more closely matched by the post-productions of the L2 learners.

Similar findings have been found for other phonetic elements as well. Accommodations of Japanese L1 speakers to the /r/-/l/ contrast in English were reliably detected by English L1 listeners (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997). Productions were recorded of the Japanese speakers both before and after perceptual identification training for the contrast. These productions were then judged by English L1 listeners, who were asked to identify which production sounded more accurate. Post-productions rated significantly higher, reflecting both the accommodation by the speakers to the perceptional (listening only) input they received, and the listener's ability to accurately detect that accommodation. A similar study by Lambacher, Martens, Kakehi, Marasinghe, and Molholt (2005) found the same results for Japanese L1 speakers' productions of specific American English vowel sounds; English L1 listeners reliably

judged the accommodation to the vowel sounds shown by the speakers after they received perceptual identification training. In social interaction, this ability to hear accommodations, even to features as subtle as the F0, could produce a feedback loop of accommodating through speech, detecting the degree of accommodation through listening, and then further adjusting the accommodation through more speech (Trofimovich, 2016).

Accommodation is also impacted by social presence. The effects are impacted by social factors such a familiarity with the interlocutor (Lelong & Bailly, 2011), and social biases (Babel, 2010). More socially interactive tasks, such as discussing possible map routes to reach a location with another individual produce accommodation which lasts longer than non-interactive word repetition (Pardo, 2006). Accommodation also occurs through more socially present environments for L2 learners, such as through exposure to an L2 dominant culture (Sancier & Fowler, 1997). Accommodation due to social contact can also be sufficient to override more conscious attitudes like national identity, as seen through historical evidence of foreign colonists phonetically accommodating to the local dialect despite a desire to maintain their original phonetic patterns (Trudgill, 2008).

The planned pair wise tests of the four conditions for the same/different discrimination task in the current study do suggest that social presence could be generated to differing degrees by the different mediation types. As discussed in sections 5.2 and 5.3 above, the videoconferencing condition displayed the most overall accuracy gains, and this was the most FTF-like condition after the videoconferencing with eye contact condition, which may have been adversely affected by creating circumstances

that could induce FLSA. These higher gains occurring in a more FTF-like condition suggest that more social presence could have been generated by the condition.

Given this, it can be argued that greater accuracy gains for detecting a single feature, such as pitch placement, reflect that accommodation has occurred. As Flege (2007) posits in the SLM, phonetic accommodation which cannot be detected through listening by an individual cannot then be consistently produced through speaking by that same individual. Learning to detect a sound feature in a language because it is necessary for comprehension and successful communication shows that phonetic accommodation is occurring; learning gains in listening detection can serve as evidence for accommodation. Therefore, the results of the current study, through the measured accuracy gains, suggest that accommodation occurred in all four mediation conditions, and further suggest that greater accommodation could occur for conditions which generate more social presence, such as videoconferencing.

5.5 Limitations of the Current Study

As mentioned throughout the preceding discussion, there are several limitations for the present study. The need for more data in each of the conditions has revealed itself in two main ways. The first is through a longer time frame for the study. Language learning is often an extended process for learners, and more time with more lessons may better emulate a language learning classroom situation. There was approximately one week between the lessons and the final testing, with no further lessons or review during that interval. In the same/different discrimination task, the two videoconferencing conditions showed greater accuracy gains in the delayed posttest, and all four conditions showed a trend of speed gains in the delayed posttest. The pattern of the speed gains does

not indicate a speed/accuracy trade-off effect, so more processing time does not seem to be needed to improve pitch placement detection accuracy after one week. Having more time with more lessons and exposures to the tutor evenly distributed throughout that time may further support and strengthen this trend, and could provide more insight into retained learning of the pitch placement distinction. This would also allow the time needed to potentially learn the vocabulary words used in the lessons, which may make the results of the picture recognition task more meaningful.

The second aspect of gathering more data would be through more participants. This would allow for smaller possible effects to be detected. For example, the interaction between mediation type and time of testing in the same/different discrimination task was tending towards significance, and this suggests that a possible significant interaction or effect for mediation type may be detectable. Also, for the correlational survey data, participants who showed accuracy loss were removed (27 removed for the posttest, 17 for the delayed posttest) so that the correlations found would only inform on potential relationships with accuracy gains facilitated by the mediation types. Because of this, the amount of data for these calculations was reduced, and more participants overall in the study would help offset this reduction.

Finally, the study has the methodological limitation of not having a FTF condition, to allow for direct comparison to the mediation types. While the general effectiveness and advantages of FTF L2 learning are well known, no claims could be made of the mediation types' performances in direct relation to FTF tutor exposure using this particular training and testing design. Although the primary purpose of the study was to compare a spectrum of mediation types in their ability to facilitate learning the pitch

placement distinction, it could be informative to also compare them directly to a FTF condition. Further study would be necessary to make this comparison and to measure the accuracy gains against FTF learning of L2 sound systems.

5.6 Implications for Second Language Acquisition and Suggestions for Future Study

Even given the limitations of the current study, some implications can be seen about the possible effects of the various mediation types used for helping language learners develop an awareness and ability to detect more subtle sound features of the L2 that are not present in the L1. All four mediation types were able to facilitate accuracy gains for pitch placement detection, showing that a range of mediation methods can be useful for this kind of awareness building. Often, subtle listening distinctions in the L2 are not a focus of classroom lessons (Foote, Trofimovich, Collins & Urzúa, 2016). However, Couper (2006) showed that short term explicit attention to specific L2 sound features in a classroom setting (12 short integrated lessons over 2 weeks) can lead to significant improvement in production in both specific and general testing. The ability to use a short-term intervention such as the one used in this study is feasible as a possible supplemental activity to provide this focused attention.

The significant time of testing effect in the same/different discrimination task shows that the lessons used were effective for teaching the distinction. The lessons used in the current study did not place an explicit emphasis through instruction on the pitch placement distinction in the Japanese language; they focused on simple sentence building. No orthographic information that would indicate pitch differences in the minimal pairs was included in the testing or lesson materials due to the study's focus on how different mediation types may facilitate learning the distinction. Yet participants

were able to realize that something subtle yet important was the cause of the minimal pairs. Several participants commented on how this made it feel like a "fun puzzle" that needed solving. This shows that lessons and tasks that aim to teach L2 grammar and vocabulary could be constructed in such a way as to also bring attention to features that are important for listening and speaking, and this can help learners attend to these features. In an actual classroom setting, explicit orthographic information could be included in such lessons; this could prove even more effective than the implicit approach used in the current study. This allows for more efficiency in lesson materials, which is often needed with limited class time.

The results also suggest that social presence is possible with these mediation types, though it is not clear to what degree for each type given this data. When taken along with the findings of Pertaub et al. (2001, 2002), who found that virtual representations of people could also generate a feeling of social presence, this further shows the promise of using mediation with audiovisual information as a language learning tool. Future research could compare these mediation types directly to a FTF learning situation, to measure the differences in accuracy gains.

There is also the possibility that these different mediation types may be more or less useful depending on the particular feature being taught. Lambacher et al. (2005) found that perceptual training improved speakers' ability to produce five specific vowel sounds in the L2, but one particular contrast between two of the vowels (/ɑ/ and /ʌ/) proved to be more difficult than the others. The authors concluded that these vowels may require more rigorous training in order to bring the needed awareness for it to be established. Different learning environments, such as different mediation types, may

better facilitate the learning of particular sound features through differing levels of emphasis on audio or visual information. Additionally, different mediation types may prove more or less useful depending on the proficiency of the learners in their L2. In the current study, all of the participants had the same proficiency level (that is, no proficiency) in Japanese, in order to control for this variable. Further comparisons of different mediation types used for various L2 proficiency levels could reveal a more detailed pattern of how the mediation types can be best applied in L2 learning.

Also, an interesting result that could be investigated with more study is a possible trend with the two synchronous conditions and retained learning. Both of the videoconferencing conditions showed higher gains for the delayed posttest, while the two non-synchronous conditions had higher gains for the posttest. This could suggest that synchronicity is more likely to lead to retained learning or stronger long-term memory formation due to the increased immediacy element of social presence it possibly generates, while lack of synchronicity results in better short-term performance. More comparisons of the presence or absence of synchronicity could also be informative for further understanding the role it plays in the generation of social presence and retained learning.

One interesting finding was that the videoconferencing with eye contact condition produced significantly less accuracy gains than the videoconferencing condition, despite providing an additional similarity to a FTF interaction. As discussed above, this could be due to the unexpected triggering of FLSA. This could suggest that for language learners with FLSA, videoconferencing without eye contact could serve as a more comfortable way to experience contact with native speakers of the L2, and may help them build

familiarity with the experience that would eventfully lead to more comfort and reduced FLSA in actual FTF encounters. The videoconferencing without eye contact condition significantly outperformed the video condition as well, which further suggests that this could serve as an ideal environment for early interactions with native speakers, as it provides social presence, but in a less socially intense manner. Çağatay (2015) also suggested that language teachers could combat the effects of FLSA in their students by exposing them more often to native speakers through the use of the internet. This coincides with the findings of this study, that suggest that videoconferencing without eye contact may be a good way to initiate these exposures. Further research could investigate this possibility, specifically by investigating the possible link between eye contact in videoconferencing and its connection to FLSA.

5.7 Conclusion

This study compared the ability of four different mediation types to facilitate the learning of a L2 sound feature, specifically the pitch placement distinction of Japanese, which can be difficult to recognize by learners whose L1 does not have the same feature. The results showed that the lessons used were effective for teaching the distinction, as all four conditions had gains in accuracy in the posttests. The results also suggested that the degree of social presence and effectiveness may differ between the mediation types, but this is unconfirmed by the current data. More data will be needed to further explore possible differences.

Future studies will need to utilize a longer study period, with more tutor exposures, participants, and conditions in order to provide a more detailed picture of the interplay between mediation, social presence, language learning, and possible FLSA

effects. Although this study showed that building effective awareness of the pitch placement distinction in Japanese is possible with minimal training and time, it only did so in one language as a case study. Future studies will be needed using different sound system features in different languages to establish the extent to which the results can be generalized. However, the current study does establish the use of mediation for building awareness and detection skills of L2 sound systems as promising for second language acquisition.

Appendix A: Surveys

Demographic Survey #_____

The following questions are for demographic information. Your responses will be completely anonymous, and will not be connected to any identifying information about you in any way. Only a participant number will be connected to this information. You may leave blank any question you do not wish to answer.

Please circle the appropriate category or fill in the requested information.

Age:

Sex: M F T

Ethnicity: Hispanic White (Non-Hispanic) Black Asian Native American Other

Handedness: Right Left

Do you have any vision or hearing problems? Yes No
If yes, what are they?:

Do you play a musical instrument? Yes No
If yes, for how long have you actively practiced/played this instrument?:

Languages spoken (next to each language listed, please write how often you use them, with "often," "sometimes," or "rarely"):

Languages you hear used around you, but that you do not speak. Next to each language listed, please write how often you hear them used, with "often," "sometimes," or "rarely." Also, write the source you hear them from, for example "TV/movies," "friends/family members," or "strangers":

Please indicate how often you use the following items or software, with 0 meaning "never", 5 meaning "monthly", and 10 meaning "daily":

Language learning CD's

0 1 2 3 4 5 6 7 8 9 10

Language learning DVD's

0 1 2 3 4 5 6 7 8 9 10

Videoconferencing programs (real-time sound and video calls, such as Skype)

0 1 2 3 4 5 6 7 8 9 10

Student Exit Survey

Part 1 Directions: Read each item carefully. Using the scale below, please circle the number that best describes the reason why you completed these lessons. Answer each item according to the following scale: 1: *corresponds not at all*; 2: *corresponds a very little*; 3: *corresponds a little*; 4: *corresponds moderately*; 5: *corresponds enough*; 6: *corresponds a lot*; 7: *corresponds exactly*.

"I completed the lessons to get course credit, but also…":

1. Because they are part of an experiment	1	2	3	4	5	6	7
2. Because I think that these lessons are interesting	1	2	3	4	5	6	7
3. Because I am doing it for my own good	1	2	3	4	5	6	7
4. Because I like to do well on all lessons	1	2	3	4	5	6	7
5. There may be good reasons to do these lessons, but personally I don't see any	1	2	3	4	5	6	7
6. Because I think that these lessons are pleasant	1	2	3	4	5	6	7
7. Because I think that these lessons are good for me	1	2	3	4	5	6	7
8. Because it is something that I have to do	1	2	3	4	5	6	7
9. I do these lessons but I am not sure if it is worth it	1	2	3	4	5	6	7
10. Because these lessons are fun	1	2	3	4	5	6	7
11. By personal decision	1	2	3	4	5	6	7
12. Because I don't have any choice	1	2	3	4	5	6	7
13. I don't know; I don't see what these lessons bring me	1	2	3	4	5	6	7
14. Because I feel good when doing these lessons	1	2	3	4	5	6	7
15. Because I believe that these lessons are important for me	1	2	3	4	5	6	7
16. Because I feel that I have to do it	1	2	3	4	5	6	7
17. I do these lessons, but I am not sure they are a good thing to do	1	2	3	4	5	6	7
18. No other reason; I completed them only for course credit	1	2	3	4	5	6	7

Part 2 Directions: Read each item carefully. Using the scale below, please circle the number that best describes the way you feel when working with a tutor. Answer each item according to the following scale: 1: *corresponds not at all*; 2: *corresponds a very little*; 3: *corresponds a little*; 4: *corresponds moderately*; 5: *corresponds enough*; 6: *corresponds a lot*; 7: *corresponds exactly*.

How do you feel when working with a tutor?

1. I feel comfortable when the tutor can see me as I work.	1	2	3	4	5	6	7
2. I feel confident when the tutor can see me as I work.	1	2	3	4	5	6	7
3. I feel anxious when the tutor watches me work.	1	2	3	4	5	6	7
4. I feel more comfortable when I can complete work without being watched by the tutor.	1	2	3	4	5	6	7
5. I like help from the tutor during my work.	1	2	3	4	5	6	7
6. I like to finish my work first, and then let the tutor check it.	1	2	3	4	5	6	7
7. I make more mistakes when doing my work in front of the tutor.	1	2	3	4	5	6	7
8. I learn better when doing my work in front of the tutor.	1	2	3	4	5	6	7

The following questions ask about future plans. Using the scale below, please circle the number that best describes your agreement with the statements. Answer each item according to the following scale: 1: *corresponds not at all*; 2: *corresponds a very little*; 3: *corresponds a little*; 4: *corresponds moderately*; 5: *corresponds enough*; 6: *corresponds a lot*; 7: *corresponds exactly*.

1. I am very likely to enroll in a course to learn Japanese language in an upcoming semester.	1	2	3	4	5	6	7
2. I am very likely to study Japanese language on my own time (self-study; not enrolled in a class) in the near future.	1	2	3	4	5	6	7

Appendix B: Test Item List

A "^" indicates a syllable with higher pitch. English translations are in brackets.

Same/different discrimination task

Practice pairs

1. "katai" [hard]; "nagai" [long]
2. "mijikai" [short]; "chiisai" [small]
3. "katai" [hard]; "katai" [hard]
4. "chiisai" [small]; "chiisai" [small]
5. "oishii" [delicious]; "ookii" [big]

Match pairs

1. "ha^shi" [chopsticks]; "ha^shi" [chopsticks]
2. "washi^" [eagle]; "washi^" [eagle]
3. "ka^ma" [sickle]; "ka^ma" [sickle]
4. "shiro^" [castle]; "shiro^" [castle]
5. "ka^ki" [oyster]; "ka^ki" [oyster]
6. "kami^" [hair]; "kami^" [hair]
7. "a^sa" [morning]; "a^sa" [morning]
8. "kame^" [pot or jar]; "kame^" [pot or jar]
9. "me^su" [scalpel]; "me^su" [scalpel]
10. "momo^" [peach]; "momo^" [peach]
11. "ki^ri" [carving tool]; "ki^ri" [carving tool]

Contrast pairs

1. "ha^shi" [chopsticks]; "hashi^" [bridge]
2. "washi^" [eagle]; "wa^shi" [traditional Japanese paper]
3. "ka^ma" [sickle]; "kama^" [iron pot]
4. "shiro^" [castle]; "shi^ro" [white]
5. "ka^ki" [oyster]; "kaki^" [persimmon]
6. "kami^" [hair]; "ka^mi" [god, deity]

7. "a^sa" [morning]; "asa^" [hemp]
8. "kame^" [pot or jar]; "ka^me" [turtle]
9. "me^su" [scalpel]; "mesu^" [female]
10. “momo^” [peach]; mo^mo” [thigh]
11. “ki^ri” [carving tool]; “kiri^” [fog]

Picture task (sound file; picture used)

Correct pairs

1. “kama^” [iron pot]; paired with correct picture
2. “hashi^” [bridge]; paired with correct picture
3. "wa^shi" [traditional Japanese paper]; paired with correct picture
4. "shi^ro" [white]; paired with correct picture
5. "kaki^" [persimmon]; paired with correct picture
6. "ka^mi" [god, deity]; paired with correct picture
7. "asa^" [hemp]; paired with correct picture
8. "ka^me" [turtle]; paired with correct picture
9. "mesu^" [female]; paired with correct picture

Incorrect pairs

10. “kama^” [iron pot]; paired with incorrect picture
11. “hashi^” [bridge]; paired with incorrect picture
12. "wa^shi" [traditional Japanese paper]; paired with incorrect picture
13. "shi^ro" [white]; paired with incorrect picture
14. "kaki^" [persimmon]; paired with incorrect picture
15. "ka^mi" [god, deity]; paired with incorrect picture
16. "asa^" [hemp]; paired with incorrect picture
17. "ka^me" [turtle]; paired with incorrect picture
18. "mesu^" [female]; paired with incorrect picture

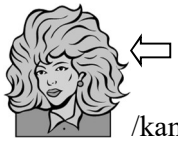
All pictures used can be seen on the Lesson 1 worksheet.

Appendix C: Lesson Worksheets

Lesson One:
Below is a list of vocabulary. Let's go over the following vocabulary, which we will use for the following lessons as well. For these lessons you only need to learn how to say the words, not how to write them in Japanese. You may use roman characters (as shown below) to write out the words. Repeat each word after you hear it.

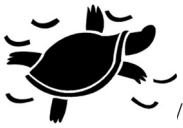

/hashi/ chopsticks

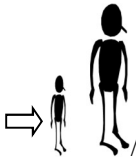

/kaki/ oyster (seafood)

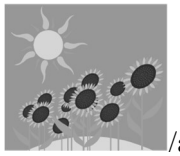

/kami/ hair


/mesu/ female


/kami/ god, deity

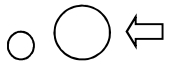

/kame/ turtle


/mijikai/ short


/asa/ morning


/oishii/ delicious


/shiro/ white (color)


/ookii/ big

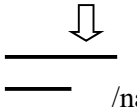

/katai/ hard (ex. stone)

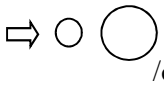

/washi/ traditional Japanese paper


/kaki/ persimmon (fruit)


/asa/ hemp (fiber/cloth)


/kama/ sickle (farm tool)


/nagai/ long


/chiisai/ small


/mesu/ scalpel (cutting tool)


/surudoi/ sharp


/kame/ pot or jar


/washi/ eagle


/shiro/ castle


/hashi/ bridge


/kama/ iron pot

(Continue to next page)

Now that we've practiced pronouncing these words, let's practice hearing them. Write the word (how to pronounce it) and English meaning of each word after you hear it.
Ex. Surudoi = sharp

1._____ 4._____

2._____ 5._____

3._____ 6._____

Now listen again. The answers will be given after each word. Check to see how you did.

Lesson Two:
Simple sentences in Japanese can express what something is. The subject is followed by the marker /wa/. The object is followed by the marker /desu/. Using these markers, simple sentences can be made using this pattern: /____wa____desu/. For example, the sentence /kama wa chiisai desu/ means "The sickle is small." Let's practice some simple sentences using the words you've learned. Repeat each sentence aloud after it is said.

1. /shiro wa ookii desu/ 'The castle is big'
2. /kame wa mesu desu/ 'The turtle is female'
3. /kaki wa oishii desu/ 'The persimmon is delicious'
4. /hashi wa chiisai desu/ 'The bridge is small'

Now let's practice completing some simple sentences. Using the words you know (you may look back on lesson one for the word list), listen to and fill in these sentences with a word you think makes sense. Say the full sentence aloud after you complete it:

1. /kame wa _____desu/
2. /_____wa surudoi desu/
3. /hashi wa _____desu/
4. /kama wa _____desu/
5. /washi wa _____desu/
6. /mesu wa _____desu/
7. /_____wa nagai desu/
8. /kaki wa _____desu/

References

Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. New York, NY: Cambridge University Press.

Arnold, P., & Hill, F. (2001). Bisensory augmentation: a speechreading advantage when speech is clearly audible and intact. *British Journal of Psychology, 92*, 339–355.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Language in Society, 39*, 437–456.

Babel, M. & Bulatov, D. (2011). The role of fundamental frequency in phonetic accommodation. *Language and Speech, 55*(2), 231–248.

Bilbatua, L., Saito, R., & Bissoonauth-Bedford, A. (2012). Using skype technology to enhance spoken language skills: Preliminary results from a pilot study in French, Japanese and Spanish at the University of Wollongong, Australia. In L.G. Chova, I.C. Torres, and A.L. Martinez (Eds.), *Edulearn12: 4th International Conference on Education and New Learning Technologies* (pp. 548-556). Valencia, Spain: International Association of Technology, Education and Development.

Bohannon, L.S., Herbert, A.M., Pelz, J.B., & Rantanen, E.M. (2013). Eye contact and video-mediated communication: A review. *Displays, 34*, 177–185.

Boiteau, T. W., Smith, C., & Almor, A. (2017). Syntax response-space biases for hands, not feet. *Attention, Perception, & Psychophysics, 79*(3), 989-999.

Bondareva, Y., Meesters, L., & Bouwhuis, D. (2006, March). Eye contact as a determinant of social presence in video communication. In M. Näel (Chair), *Session 7: Multimodal, multi-device, context dependent systems, services and applications.* Symposium conducted at the 20th International Symposium on Human Factors in Telecommunication, Sophia-Antipolis, France.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America 101*, 2299-2310. doi: 10.1121/1.418276

Buxton, B., & Moran, T. (1990). EuroPARC's Integrated Interactive Intermedia Facility (IIIF): Early Experiences. In S. Gibbs, & A. A. Verrijn-Stuart (Eds.), *Proceedings of the IFIP WG 8.4 Conference on Multi-user Interfaces and Applications* (pp. 11-34). Amsterdam, Netherlands: Elsevier Science Publishers, B.V.

Çağatay, S. (2015). Examining EFL students' foreign language speaking anxiety: The case at a Turkish state university. *Procedia - Social and Behavioral Sciences, 199*, 648–656.

Cho, M-H., & Heron, M. L. (2015). Self-regulated learning: The role of motivation, emotion, and use of learning strategies in students' learning experiences in a self-paced online mathematics course. *Distance Education, 36*(1), 80-99. doi: 10.1080/01587919.2015.1019963

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect, 21*(1), 46-66.

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics, 28*, 411-424.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York, NY: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268. doi: 10.1207/S15327965PLI1104_01

DeLoache, J.S., Chiong, C., Sherman, K., Islam, N., Vanderborght, M., Troseth, G.L., Strouse, G.A., & O'Doherty, K. (2010). Do babies learn from baby media? *Psychological Science, 21*(11), 1570-1574.

Doherty-Sneddon, G., Anderson, A., O'Malley, C., Langton, S., Garrond, S., & Bruce, V. (1997). Face-to-face and video-mediated communication: A comparison of dialogue structure and task performance. *Journal of Experimental Psychology: Applied 3*, 105–125.

Ducate, L., & Arnold, N. (2011). Technology, CALL, and the Net Generation. Where are we headed from here? In L. Ducate and N. Arnold (Eds.) *Present and Future Promises of CALL: From Theory and Research to New Directions in Language Teaching.* (pp. 1-22). San Marcos, Texas: CALICO.

Dupont, S., Aubin, J., & Menard, L. (2005). A study of the McGurk effect in 4- and 5-year-old French Canadian children. *ZAS Papers in Linguistics, 40*, 1–17.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole, & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353-381). New York, NY: Mouton de Gruyter.

Foote, J. A., Trofimovich, P., Collins, L., & Urzúa, F. S. (2016). Pronunciation teaching practices in communicative second language classes. *The Language Learning Journal, 44*(2), 181-196. doi: 10.1080/09571736.2013.784345

Forster, K.I. & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35* (1), 116-124.

Fullwood, C., & Doherty-Sneddon, G. (2006). Effect of gazing at the camera during a video link on recall. *Applied Ergonomics, 37*, 167–175.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers.* New York, NY: Academic Press.

Gregory, S.W., Green B.E., Carrothers, R.M., Dagan, K.A. & Webster, S.W. (2001). Verifying the primacy of voice fundamental frequency in social status accommodation. *Language & Communication, 21*, 37-60.

Guay, F., Vallerand, R.J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (SIMS). *Motivation and Emotion, 24*(3), 175-213.

Guichon, N., & Cohen, C. (2014). The impact of the webcam on an online L2 interaction. *The Canadian Modern Language Review / La Revue canadienne des langues vivantes, 70*(3), 331-354.

Gumperz, J. J. (1982). Contextualization conventions. In *Discourse strategies* (pp. 130-152). New York, NY: Cambridge University Press.

Gumperz, J. J. (2003). Interactional Sociolinguistics: A Personal Perspective. In D. Schiffrin, D. Tannen, & H.E. Hamilton (Eds.), *The Handbook of Discourse Analysis*. Blackwell Reference Online. doi: 10.1111/b.9780631205968.2003.00012.x

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132.

Jauregi, K., de Graaff, R., van den Bergh, H., & Kriz, M. (2012). Native/non-native speaker interactions through video-web communication: A clue for enhancing motivation? *Computer Assisted Language Learning, 25*(1), 1-19.

Kuhl, P. K., Tsao, F.-M. & Liu, H.-M. (2003). Foreign language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, 100*(15), 9096–9101.

Lambacher, S., Martens, W., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*, 227–247. doi: 10.1017.S0142716405050150

Lawson, T., Comber, C., Gage, J., & Cullum-Hanshaw, A. (2010). Images of the future for education? Videoconferencing: A literature review. *Technology, Pedagogy and Education, 19*(3), 295-314.

Lee, L. (2007). Fostering second language oral communication through constructivist interaction in desktop videoconferencing. *Foreign Language Annals, 40*(4), 635-649.

Lelong, A., & Bailly, G. (2011). Study of the phenomenon of phonetic convergence thanks to speech dominoes. In A. Esposito, A. Vinciarelli, K. Vicsi, C. Pelachaud, & A. Nihjolt (Eds.), *Analysis of verbal and nonverbal communication and enactment: The processing issue* (pp. 280–293). New York: Springer.

Levinson, S. C. (2003). Contextualizing "contextualization cues". In S.L. Eerdmans, C.L. Prevignano & P.J. Thibault (Eds.), *Language and Interaction: Discussions with John J. Gumperz* (pp. 31-39). Philadelphia, PA: John Benjamins.

Massaro, D. W., Thompson, L., Barron, B., & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. *Journal of Experimental Child Psychology, 41*, 93–113.

McDonough, K., Crowther, D., Kielstra, P., & Trofimovich, P. (2015). Exploring the potential relationship between eye gaze and English L2 speakers' responses to recasts. *Second Language Research, 31*(4), 563-575.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746–748.

Morton, H., & Jack, M. (2010). Speech interactive computer assisted language learning: A cross-cultural evaluation. *Computer Assisted Language Learning, 23*, 295–319.

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal of Language and Social Psychology, 21*, 422–432.

Navarra, J., & Soto-Faraco, S. (2007). Hearing lips in a second language: visual articulatory information enables the perception of second language sounds. *Psychological Research, 71*, 4-12.

Neureiter, K., Fuchsberger, V., Murer, M., & Tscheligi, M. (2013). Hands and eyes: How eye contact is linked to gestures in video conferencing. In W.E. Mackay, S.A. Brewster, & S. Bødker (Eds.) *Proceedings of CHI '13 Extended Abstracts on Human Factors in Computing Systems* (pp. 127-132). New York: Association for Computing Machinery.

Nielson, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*, 132–142.

Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*(4), 2382–2393.

Pardo, J.S., Gibbons, R., Suppes, A. & Krauss, R.M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*, 190–197.

Pardo, J.S., Jordan, K., Mallari, R., Scanlon, C. & Lewandowski, E. (2013). Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures. *Journal of Memory and Language, 69*, 183–195.

Pearson, B.Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics, 28*, 399-410.

Pertaub, D-P., Slater, M., & Barker, C. (2001) An experiment on fear of public speaking in virtual reality. In J. D. Westwood, H. M. Hoffman, G. T. Mogel, D. Stredney, & R. A. Robb (Eds.), *Medicine Meets Virtual Reality 2001: Outer Space, Inner Space, Virtual Space* (pp. 372-378). Amsterdam, Netherlands: IOS Press.

Pertaub, D-P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence: Teleoperators and Virtual Environments, 11*(1), 68-78.

Pierrehumbert, J. & Beckman, M. (1988). *Japanese Tone Structure*. Cambridge, Massachusetts: MIT Press.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–190.

Poser, W. (1984). *The phonetics and phonology of tone and intonation in Japanese* (Doctoral dissertation). Retrieved from http://hdl.handle.net/1721.1/15169

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*, 421–436.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications.* London, UK: John Wiley & Sons.

Sugito, M. (1998). *Nihongo onsei no kenkyu 5: "hana" to "hana"* [Studies on Japanese Phonetics 5]. Osaka, Japan: Izumi Shoin.

Sugiyama, Y. (2006). Japanese pitch accent: examination of final-accented and unaccented minimal pairs. *Toronto Working Papers in Linguistics, 26*, 73–88.

Trofimovich, P. (2013). Interactive alignment: Implications for the teaching and learning of second language pronunciation. In J. Levis & K. LeVelle (Eds.). *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference.* (pp. 1-9). Ames, IA: Iowa State University.

Trofimovich, P. (2016). Interactive alignment: A teaching-friendly view of second language pronunciation learning. *Language Teaching, 49*(3), 411–422. doi:10.1017/S0261444813000360

Trudgill, P. (2008). Colonial dialect contact in the history of European languages: On the irrelevance of identity to new-dialect formation. *Language in Society, 37*, 241–280.

Vance, T. (1995). Final accent vs. no accent: utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics 23*, 487–499.

Wang, Y. (2001). *Perception, production, and neurophysiological processing of lexical tone by native and non-native speakers* (Doctoral dissertation). Ann Arbor, MI: Proquest.

Wang, Y., & Chen, N-S. (2012). The collaborative language learning attributes of cyber face-to face interaction: The perspectives of the learner. *Interactive Learning Environments, 20*(4), 311-330.

Wayland, R.P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36*, 250–267.

Winters, S. & Grantham O'Brien, M. (2013). Perceived accentedness and intelligibility: The relative contributions of F0 and duration. *Speech Communication, 55*, 486–507.

Yamada, M. (2009). The role of social presence in learner-centered communicative language learning using synchronous computer-mediated communication: Experimental study. *Computers & Education, 52*, 820-833.

Yamada, M., & Akahori, K. (2007). Social presence in synchronous CMC based language learning: How does it affect the productive performance and consciousness of learning objectives? *Computer Assisted Language Learning, 20*(1), 37-65.

Yanguas, Í. (2010). Oral computer-mediated interaction between L2 learners: It's about time! *Language Learning & Technology, 14*(3), 72-93.

Yip, B., & Jin, J. S. (2003). An effective eye gaze correction operation for video conference using anti-rotation formulas. In *Proceedings of the Joint Conference of the 4th International Conference on Information, Communications and Signal*

Processing/4th Pacific-Rim Conference on Multimedia: ICICS-PCM 2003 (pp. 699-703). doi: 10.1109/ICICS.2003.1292546

Zhan, Z., & Mei, H. (2013). Academic self-concept and social presence in face-to-face and online learning: Perceptions and effects on students' learning achievement and satisfaction across environments. *Computers & Education, 69*, 131-138.