Hand Images in Virtual Spatial Collaboration for Map-Based Planning Activities

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Claire Xuan Cheng

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

Co-Advisers: Caroline Hayes, Renata Wentzcovitch

October 2014

Acknowledgements

I would like to express my gratitude to my supervisor Caroline Hayes for the useful comments, remarks, and engagement through the learning process of this master thesis. I would like to thank you for encouraging my research and for allowing me to grow in my writing ability. Your advice on both research and writing skills has been priceless. With special gratitude, I thank my program director and co-adviser, Renata Wentzcovitch, whose contribution in stimulating suggestions and encouragement has helped me to coordinate my thesis and put things together. I would also like to thank Professor Massoud Amin for serving as my committee member. I want to thank all the committee members for letting my defense be an enjoyable moment and for your brilliant comments and suggestions. Thank you.

Also, I would like to thank the participants in the user-experience study for this project, who have all willingly shared their precious time during the process of user testing and interviewing. I also want to thank my colleagues who were involved in this project for their contributions—Andrew Ahlfield, Daniel Drew, Mai Anh Nguyen, and Ashley Clayson. Thank you for your brilliant ideas and hard work, continually inspiring me. I would also like to thank all the friends who supported me in writing and encouraged me to strive towards my goal.

Abstract

The goal of this project is to improve understanding about the communication channels that assist distant collaborators to perform more effectively when collaborating in a virtual environment. The motivation is to help software developers to decide on the features that should be included in virtual collaboration tools. This work focuses on communication through voice, gestures conveyed via natural hand images, shared maps, markings on maps, and combinations of the above. The task domain studied includes joint, map-based planning tasks, which range from trip planning to traffic disaster management, such as a truck rollover on a high way. **Embedded natural gestures** are made with the hands or body and derive a meaning from their context, such as, a person pointing to a location on a map; in this work, we will refer to them simply as natural gestures. **Surrogate gestures** are electronic proxies for natural gestures and include pointing with a cursor or drawing circles, arrows, and other marks on the map. Both natural and surrogate gestures are major concerns in this work. Currently, remote collaborations between traffic experts at different agencies (for example, the state and the city) are usually carried out telephonically. Over the past twenty-five years, new tools have been developed that allow collaborators to work in a shared virtual workspace in which they can not only see shared images and mark shared drawings, but they also see the hands of their distant partners as they move over the work surface. However, few researchers have evaluated the effectiveness thereof. The primary questions explored in this work are whether embedded natural gestures or surrogate gestures provide significant advantages over voice-only communication in virtual collaborations regarding map-based tasks. The answers to these questions could help software developers decide on the features to include in virtual collaboration tools. In order to answer these questions, we recruited twenty-eight students, both undergraduate and graduate, to participate in an experiment. The participants worked in pairs to solve five map-based planning tasks using five versions of map-based workspaces. These five versions of workspaces were created by combining different interface features that supported diverse types of communication: voice, a shared virtual-map interface, a shared marking interface (to support surrogate gestures), and a hand-image interface (to support natural gestures). We set up five different combinations of interfaces, as follows:

- **Face-to-Face**: Collaborators sitting side-by-side share a virtual-map workspace on which they can both make marks;
- **Voice-only**: Distant collaborators can manipulate and mark separate virtual-map workspaces, but cannot share workspaces, and can only communicate vocally;
- **Mark-Voice**: Distant collaborators have a shared virtual-map workspace on which they can mark and share marks, and can also communicate vocally;
- **Gesture-Voice**: Distant collaborators have a shared virtual-map workspace in which they can see videos of each other's hands and arms projected on the map, and also communicate vocally**;**
- **Mark-Gesture-Voice**: Distant collaborators have a shared virtual-map workspace, on which they can see each other's marks and gestures, and they can communicate vocally.

The pairing of the interface conditions and task scenarios was systematically varied so that the same interface condition and task scenario were not always paired together. In addition, the presentation order was systematically varied. After each condition, we asked each of the participants six questions about their workload from the NASA Task Load Index and seven questions about their collaborative experience. We found that

- From performance perspective, all the conditions that involved using embedded natural gestures (Gesture-Voice, and Mark-Gesture-Voice) significantly 1) reduced task completion time, 2) decreased mental demand and 3) helped participants felt more connected to their teammates; additionally, when using the Gesture-Voice condition, participants experienced significantly less frustration and collaborated significantly more seamlessly than in the Voice-Only condition.
- From preference perspective, Mark-Gesture-Voice was 1) the easiest to use, 2) the most fun, 3) the mostly chosen as professional collaboration tools, 4) the one that helped the user felt like most connected with their partners among all the remote conditions and 5) the favorite among all the remote conditions; even though participants still like the Face-to-Face condition better than any of remote conditions and felt it easiest to use among all the conditions.

We can, then, conclude that the hand images are the element primarily responsible for the performance improvement in remote collaboration, but that users enjoy having the marking feature, regardless of

whether it helps them significantly or not. Based on these findings, we recommend that software developers of virtual-collaboration tools should include hand images to improve performance, and should also consider including a shared-marking function to increase user-satisfaction.

Table of Contents

List of Tables

List of Figures

1. Introduction

The goal of this project is to better understand which communication channels could help distant collaborators to perform better when collaborating in a virtual environment. An understanding of this question will help software developers decide which features to include in the virtual collaboration tools. The task domain in this work is *gestures*, both natural and surrogate. **Embedded natural gestures** are made with hands or body and derive a meaning from their context, for example, the gesture made by a person pointing to a location on a map. **Surrogate gestures** are electronic proxies for natural gestures; examples of this include drawing circles, arrows, and other operations on an electronic map by pointing on the screen with a cursor.

1.1 Motivation

This section will address the following questions: 1) What problem is explored in this thesis and what are the existing collaborative tools for this domain? 2) What are the features that best support virtual collaboration for this domain? 3) On what prior work is the current project based?

Many companies and researchers have created applications for virtual teams that bring together geographically dispersed workers with complementary skills [8, 13, 29, 31]. These tools explore a wide range of features to support virtual collaborations, including shared voice, video, drawing, gestures, or teleoperated robots. However, relatively little work has evaluated the effectiveness of those tools in supporting virtual collaboration. This thesis focuses on assessing the effectiveness of several features in a virtual collaboration tool for supporting map-based tasks. These features include shared virtual maps, shared markings on the maps, and the ability to view each other's natural gestures. In this work, we chose to focus on natural and surrogate gestures for several reasons. Gestures appear to be very important in communication. Natural gestures are often lacking in distance collaboration because they could require significant bandwidth and extra processing over surrogate gestures. Do natural gestures provide enough advantages to merit the expenditure of extra resources? Do either natural or surrogate gestures provide sufficient advantages over old-fashioned, but reliable, collaboration by telephones? Experimental studies [23, 46] have indicated that simply linking remote spaces through audio-visual video links (as opposed to audio-only) does not improve collaborative performance, especially not to the levels observed in co-present collaboration [3, 4]. Fussell and her colleagues have argued that to facilitate the performance in completing the remote collaborative tasks, a representation of gestures between the virtual spaces should be provided [8]. Also, the research suggests that shared visual access to collaborative task spaces is effective in the establishment of conversational grounding (i.e., the establishing of mutual knowledge, beliefs, attitudes, and expectations) [17, 18, 19]. Additionally, gestures are important in establishing common understandings [18, 19]. Therefore, the focus here is on support of gesture use in virtual collaboration.

This study is directly based on work by both Caroline Cornelius [36] and Daniel Drew [10], and is a direct follow-up to Drew's project on hand images in virtual collaboration for traffic incidents and disaster management. Cornelius' project assessed whether adding hand images to sketching could provide more benefit than sketching only, within the area of collaborative, conceptual product design. Her results showed that users performed significantly better with hand images and sketching than with sketching only. Then, Drew (with the assistance of the author) developed a test bed and an experiment to investigate the impact of hand images in virtual collaboration [10] for traffic incident and disaster management. However, his results did not show a clear advantage for adding hand images to markings on a shared virtual map, beyond simply allowing participants to make markings on a shared, virtual-map workspace. Such a result was very different than that found by Cornelius [37], who found that in a shared, virtual workspace, gestures plus marking reduced workload significantly over marking by itself. Several extensions to Drew's work may further clarify his results. His experimental conditions did not examine a hand image condition without marking, thus the impact of hand images independent of marking is unknown; furthermore, the experimental conditions did not compare against the base-line of Voice-only (e.g., communication through phones only) that represents the way that traffic management problems are typically carried out today.

Thus, in order to clarify the impact of gestures in virtual collaboration, the new experiment is a modified version of Drew's experiment, which also isolated the hand-image variable and provided an additional, realistic baseline condition: Voice-only (by phone). The current project hypothesizes that providing distant collaborators with images of their partners' hands will reduce task-completion time, decrease their workload, and increase team cohesion. In particular, this project specifically explored the impact of hand images on: 1) the performance as measured by task-completion time; 2) the remote collaborators' workload as measured by the NASA Task Load Index (a subjective, multidimensional assessment tool that rates perceived workload in order to assess a task, system, or team's effectiveness or other aspects of performance); and 3) the team cohesion as measured by the feeling of connection to the collaborators and the seamlessness of work as a team, despite the dispersal across multiple locations. This project provides the scientific understanding necessary to inform the development of more effective tools to support experts in designing map-based planning activities.

1.2 Approach

The thesis question was explored through the following steps: 1) modifying the experimental test-bed originally used in Cornelius' project [37], and adapted the software that I developed and improved in Drew's project; 2) running the experiment under five different task scenarios and conditions (including the additional Gesture-Voice condition and Voice-only condition), and then I collected data through experiments; and 3) analyzing the data and drew conclusions.

Firstly, according to the features of distance collaboration in map-based planning activities, this work built the workspace based on Cornelius' virtual collaboration project [37], implementing the same hardware setup. However, the original sketching collaborative interface from Cornelius's project could not be used for map-based planning activities, so new functions were developed for the current project and parts of the software were updated, based on Cornelius' project, ensuring a significantly better performance. New functions consist of the shared virtual-map interface and the shared marking interface. **The shared virtual-map interface** allows two people at different locations to view the same virtual map as projected on a table top. They can drag, zoom in and out of the virtual map and enable or disable the map's traffic view. **The shared marking interface** allows two people at a distance to share their drawings and markings on the shared virtual-map interface. They share those virtual-map and -marking interfaces, but have separate cursors, both controlled by infrared pens. The updated software is **the shared hand-image software**, which can project the distance collaborator's hand and arm images on the top of the local collaborator's virtual workspace. Through my improvement, the noise of the hand videos, present in the old software, significantly decreased and the projections of the hand videos through the current software are much smoother.

Secondly, after the development of the whole software platform, we recruited twenty-eight undergraduate and graduate students to participate in an experiment. In the experiment, participants worked in pairs using five different versions of a map-based virtual workspace to jointly solve the map-based planning problems. The five conditions of workspaces are:

- **Face-to-Face**: Two participants sit on the same side of the wall using one virtual-map workspace. They can both mark on the shared virtual-map workspace.
- **Voice-only**: Two participants sit on opposite sides of a wall using separated virtual-map workspaces. They can mark on their own virtual maps, but they cannot share their markings. They can only communicate through voice.
- **Mark-Voice**: Two participants sit on opposite sides of a wall using distributed shared virtual-map workspaces. They can mark on their virtual maps and share their markings. They can also communicate through voice.

- **Gesture-Voice**: Two participants sit on opposite sides of a wall using separated virtual-map workspaces. They can feel each other's presence by seeing the hand and arm images of their partner on their virtual maps. They can also communicate through voice.
- **Mark-Gesture-Voice**: Two participants sit on opposite sides of a wall using distributed shared virtual-map workspaces. They can mark on their virtual maps, share their markings and feel each other's presence by seeing the hand and arm images of their partner on the virtual maps. They can also communicate through voice.

We ran the experiments under the above five conditions with five different task scenarios, using a mix-model design. The mix-model takes individual differences of interface condition into account while computing the fixed effects. During the experiments, we recorded the time that participants spent in each task under different interface conditions. After each interface condition, the participants were asked to fill out a post-task questionnaire based on the NASA Task Load Index and the participants' remote collaborative experience; also, after a series of five conditions, the participants were asked to fill out a concluding questionnaire based on their preferences. When the whole experiment was finished, we interviewed the participants based on their questionnaire results and audio-recorded the interviews. Then, we collected data for the completion times, questionnaire results, and interview records.

Lastly, using those collected data sets, we compared the completion times under different scenarios, analyzed the questionnaire results using various statistical analyses, and recorded the interviews. Those evaluation methods are based on the previous projects from the same lab (Cornelius' project [37] and Drew's project [11]). We found that,

- From performance perspective, Gesture-Voice helped the most often, followed by Mark-Gesture-Voice:
    - Compared with Voice-Only, the conditions that involved the gesture projection (Gesture-Voice, and Mark-Gesture-Voice) significantly
        1) Reduced task completion time,
        2) Decreased mental demand,

3) Helped participants feel more connected to their teammates;

    - Additionally, compared with Voice-only, when using the Gesture-Voice condition, participants

        1) Felt significantly less frustrated,

        2) Collaborated significantly more seamlessly.

- From preference perspective, Face-to-Face is the favorite, followed by Mark-Gesture-Voice:

    - Among all the remote conditions, the Mark-Gesture-Voice condition was the

        1) Easiest to use,

        2) Most fun,

        3) Most often identified as their top choice as a professional collaboration tool,

        4) One that helped the user feel most connected with their partners,

        5) Favorite;

    - Additionally, participants still like the Face-to-Face condition better than any of remote conditions and felt it easiest to use among all the conditions.

Thus, we concluded that the hand images are significantly helpful in map-based distance collaboration tasks. Based on these findings, we recommend that developers of virtual collaboration tools should include hand images as a method of support. They should also consider including a marking function to allow the users to mark on the map, because it makes them happy and does not hurt performance.

2. Contributions

My thesis contributions have been in two related projects, each with multiple collaborators, and with a focus on the latter:

- “Hand Images in Virtual Spatial Collaboration for Traffic Accident and Disaster Management” [10]. Daniel Drew led this project.
- “Hand Images in Virtual Spatial Collaboration for Map-Based Planning Activities.” I led this project jointly with Andrew Ahlfield.

In the first project, “Hand images in Virtual Spatial Collaboration for Traffic Accident and Disaster Management” [10], led by Daniel Drew:

- I developed a shared virtual-map interface that helped two individual distance collaborators to jointly manage traffic accidents and disasters;
- I developed a shared marking interface to help distance collaborators share their markings on the virtual map;
- I improved the hand-images software from Cornelius’ project and applied the improved software to measure the effects of hand images in virtual distance collaboration, for which the new hand-image software renders hand images smoothly in real time, with no eco-noise;
- I was part of the analysis team, composed of three coders, to code the experiments’ videos. There were a total of fifty-four coding tasks (six coding tasks for each pair; nine pairs of participants in all). Eighteen of the fifty-four video coding tasks were completed by all three coders to ensure that the interpretations were correct. The three coders each represented different functions: a software developer (me), an experimenter (Daniel Drew), and an outside party (Ashley Clayson). The remaining thirty-six tasks were divided evenly among the three coders. Thus, each coder coded thirty video-coding tasks.

In the second, current project: "**Hand images in Virtual Spatial Collaboration for Map-Based Planning Activities,**" led by me, jointly with Andrew Ahlfield:

- I improved the marking interface from the first project (led by Drew [10]) in two aspects—the "Making Tools" buttons were enlarged for user-friendly purposes and an "Erase All" confirmation window was added to prevent the participants from accidentally deleting all the markings.
- I jointly completed the experimental design with Andrew Ahlfield. We discussed with each other the task and questionnaire design to develop the final plan for the experiment. The goal of the task design was to maintain variety and to control complexity. So each of us designed five scenarios and then by agreement, we jointly decided which five we would use in the current project: the final five task scenarios contain two suggestions from me (the Hawaii Boating Excursion and the Mall Run task), two suggestions from Andrew Ahlfield (the Bicycle Trip and Hiking Trail task), and one from Daniel Drew's experiment (the Bomb Threat/Evacuation task).
- I concurrently ran the experiments with Andrew Ahlfield. There were fourteen series of experiments in all. Andrew Ahlfield and I completed eight of fourteen jointly. Both of us were present at the beginning of each of those eight, so as to be better able to address problems. But once the experiments were running smoothly, Andrew Ahlfield and I separately controlled the remaining six experiments. Consequently, each of us ran a total of eleven experiments.
- I jointly interviewed all the participants with Andrew Ahlfield for ten pairs of real tests and eight pairs of beta tests, twenty-eight participants in all (each pair of test was completed by two participants).
- I completed the results analysis separately from Andrew Ahlfield.
    1) Each of us independently coded the videos to determine the task-completion time for ten pairs of participants, fifty tasks in all (each pair of tests included five tasks). This double coding is essential to ensure the correctness of interpretation. The times' difference we identified was less than 10 seconds. Therefore, we chose the average time as the task-completion time.

2) Each of us independently analyzed the task-completion times and all the questionnaire results using ANOVA analysis, pairwise comparison, and mean analysis. Ten pairs of real tests, 120 task-completion times, and 120 questionnaire results in all (each pair of subjects completed one post-task questionnaire after each task of five and also completed one concluding questionnaire after all five tasks). **My focus in the data analysis is ANOVA analysis, pairwise comparison, and mean comparison, while Andrew Ahlfield's focus was video coding.**

The remainder of this thesis will be focused on the second project.

3. Literature Review

Remote Gesture Technologies have been developed based on the relative perceived role of gestures in the activity to be supported. These kinds of tools treat the representation of gestures as an important communication tool in each of the application areas, for example, medical support applications (Figure 1), remote collaborative physical tasks' applications (Section 3.1), and remote collaborative mixed-presence applications (Section 3.2). The following sections will present a review of the existing remote gesture technologies in selected areas.

Figure 1: An example of gestural communication over a video stream in a medical context [13]. Reprinted Courtesy of the Association ACM for Non-Commercial Machinery.

3.1 Remote Gesture Technology to Support Physical Tasks Collaboration

The first broad category of remote gesture systems is the support of collaborative physical tasks. A collaborative physical task is a class of tasks in which two or more people work together with physical objects in the real world and which have an inherent 2D or 3D structure (for example, completing a 2D-robot draft design or making a 3D robot) [36]. These tasks are usually focused on "instruction"[8]: one person (whom they call the "worker") directly manipulates objects and tools under the guidance of another person (whom they call the "helper"), who provides instructions but does not physically manipulate objects. In these tools, the representation of gestures is an artifact of communication used to facilitate ongoing remote collaborations.

3.1.1 Drawing Gestures over Video Environment System

The *Drawing Gestures over Video Environment (DOVE) System* is a system integrating drawing gestures and live video to support collaboration on physical tasks, led by Susan R. Fussell [8]. Their study considered tools to support remote gestures in video systems being used to complete collaborative physical tasks under a helper—worker structure: the helper helped the worker fix a small robot, as shown in Figure 2.

The creation of the DOVE system was motivated by discussions about what took place in the process of conversational grounding during collaborative physical tasks, particularly by discussions about the role of two types of gestures in the grounding process: ***pointing gestures***, which are used to refer to task objects and locations, and ***representational gestures***, which are used to represent the form of task objects and the nature of actions to be used with those objects. Thus, Fussell and her colleagues considered ways in which both pointing and representational gestures can be instantiated in systems for remote collaboration on physical tasks [8]. Their studies used a "surrogate" approach to remote gesturing, in which images are intended to express the meaning of gestures through visible embodiments, rather than direct views of the hands.

Figure 2: Close-up of the DOVE drawing tool on the helper's tablet PC (left front insert) and on the worker's monitor (right) [8].

The DOVE system's architecture combines network IP cameras, desktop PCs, and tablet PCs (shown in Figure 3) to allow a remote helper to draw on a video feed of a workspace, as he or she provides task instructions. Under this architecture, the workspace is equipped with tablet PCs, desktop PCs, or other handheld devices that visually interact through video cameras. Additionally, a gesture-recognition component enables the system both to normalize freehand drawings in order to facilitate communication with remote partners and also to use pen-based input as a camera control device. Real-time video streams from these cameras are sent to collaborators' computing devices in the workspace. Thus, a helper can make freehand drawings and pen-based gestures on the touch sensitive screen of a computing device that has been overlaid on the video stream (shown as the front-inset picture of Figure 2). The resulting mixed image is presented back to the workers via a monitor in their local task space (shown as the background picture of Figure 2). The results are observable by all collaborators on their own monitors.

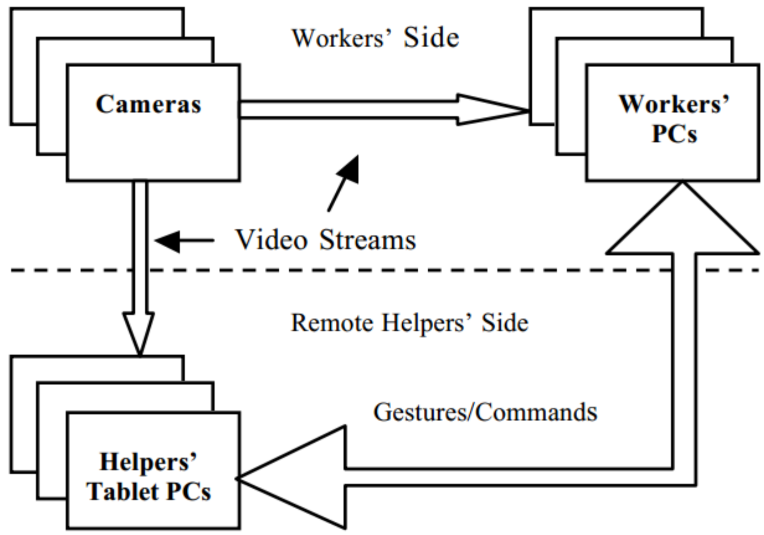


Figure 3: Overview of the DOVE system architecture [13]. Reprinted Courtesy of the Association ACM for Non-Commercial Machinery.

Fussell and her colleagues' study included two parts. In the Study 1 of ***pointing gestures***, they compared performance with a cursor-based pointing device that allows remote partners to point to objects in a video feed of the work area. Forty-eight pairs of undergraduates completed three robot assembly tasks, one in each of three media conditions:

- **Video-only**: The helper could view the output of the camera focused on the worker's task environment, but could not manipulate the video feed;
- **Side-by-side**: The helper and worker worked side-by-side;
- **Video + cursor pointer**: The helper could point to objects in the video feed.

The evaluation resources that Fussell and her co-authors used in Study 1 were task-completion times and post-task questionnaire responses. From the statistical analysis, they found the following results:

- **Task-completion time (**referred to as "performance"**)**: Adding a cursor pointer to the video system was not sufficient to improve performance over that in the video-only condition; performance was only significantly faster in the side-by-side condition compared to the two other conditions.

- **Post-Task Questionnaire** (including coordination questions and ease of identifying referent questions):
    1) **The Coordination** under the video + cursor condition and the side-by-side condition were all significantly higher than the video-only condition;
    2) **The Ease of Identifying Referents'** questions and responses showed significant differences between all the conditions; the side-by-side condition was the easiest to identify referents and the video-only condition was the hardest.

Based on their analyses in Study 1, Fussell and her colleagues concluded that although participants reported finding value for the cursor-pointing device from the post-task questionnaire responses, the tool did not improve performance times over the video alone.

In the Study 2 of ***representational gestures***, Fussell and her colleagues compared performance with two variations of a pen-based drawing tool that allows for both pointing and representational gestures (as shown in Figure 4) in a video feed of the work area. Twenty-eight pairs of undergraduate students completed three robot assemblies, one in each of three media conditions:

- **Video-only**: The helper could view the output of the camera focused on the worker's task environment, but could not manipulate the video feed;
- **DOVE + Manual-erase**: The helper could draw on the video feed but had to manually erase the drawing gestures;
- **DOVE + Auto-erase**: The helper could draw on the video feed and the drawing gestures disappeared after three seconds.

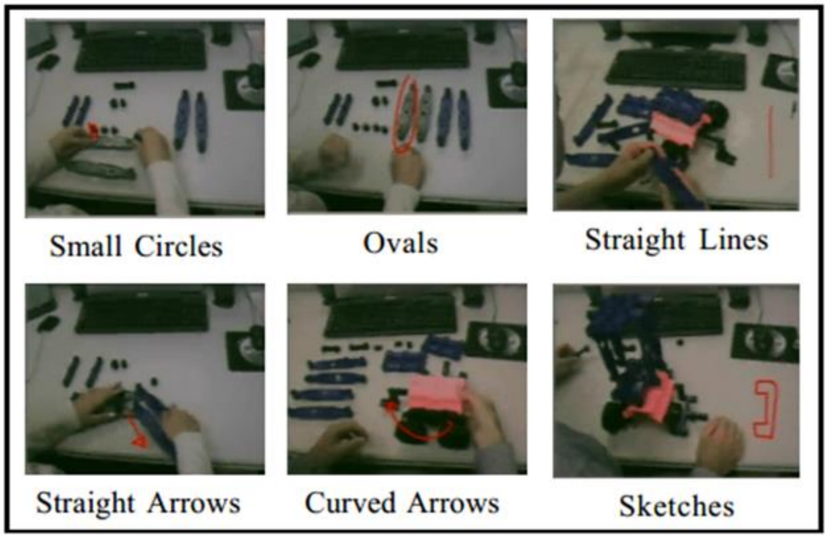


Figure 4: Sample of representational gestures created by participants using DOVE during a robot assembly task [13]. Reprinted Courtesy of the Association ACM for Non-Commercial Machinery.

The evaluation resources used in Study 2 of the DOVE system are task-completion times, post-task questionnaire responses, and final questionnaire responses, which were answered after all three tasks were completed to rate the value of the technology's features; in the final questionnaire, both helpers and workers rated both versions of the drawing software as "helpful" and "preferred" for their collaboration. From the statistical analysis, the following results were found:

- **Task-completion time**: The difference between the manual-erase and the video-only condition was not significant; the auto-erase condition was significantly faster than both the manual-erase and the video-only conditions.
- **Post-Task Questionnaire**:
    1) **The Coordination** under the manual-erase condition and the auto-erase condition were all significantly higher than the video-only condition;
    2) **The Ease of Identifying Referents'** questions and responses showed that the manual-erase and the auto-erase conditions were all significantly easier than the video-only condition.
- **Final Questionnaire**: No difference was found between the ratings of the two types of erasure regarding helpfulness; helpers showed a slight preference for the auto-erase mode.

Based on the above analysis of Study 2, Fussell and her colleagues concluded that a pen-based drawing tool could facilitate task communication and performance on their collaborative robot-construction task, particularly when the drawings had been automatically erased after several seconds; performance times with the auto-erase version of DOVE were nearly identical to those reported for pairs working side-by-side; in addition, conversations using DOVE in auto-erase mode were more efficient than those using manual-erase mode or video-only.

In conclusion, the simple, surrogate-gesture tools, no matter whether tools of pointing or representational gestures, did not significantly differ from what was found from the video-only condition. Only when an automatic-erasure function, in which drawings disappeared a few seconds after they were created, was added to the representational gesture tools did the remote collaboration's performance significantly improve from the video-only condition. The conclusion of my study, that representational gestures alone did not create a significant improvement, is consistent with findings from the studies of Kirk and Fraser [16], Cornelius [37], and Drew [10].

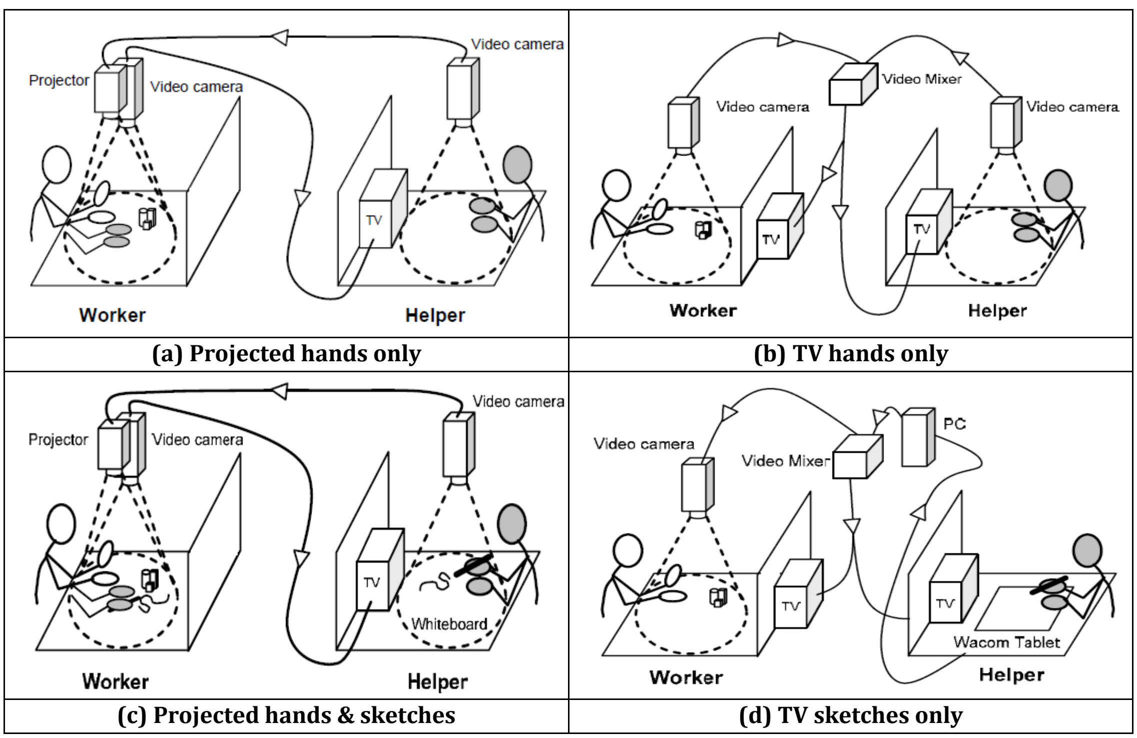
3.1.2 Remote Gesture Technologies

David S. Kirk and Danaë Stanton Fraser's [16] study comparing remote gestures was applied here to support remote collaborations on physical tasks, also. Different from the DOVE system, Kirk and Fraser's project added the vehicles for the direct view of hands. The motivation of their study was that systems for remote gesture technology had taken divergent approaches to the representation of remote gestures. Those other systems had been constructed using remote gesture systems with combinations of three different gesture formats (hands only, hands and sketch, and sketch only) and two different gesture output locations (projected hands—directly projected onto a worker's task space without mediation, or TV hands—presented through a video using a TV). There was no clarity about which approach had the most benefit for task performance. So Kirk and Fraser's project was aimed at finding out which approach to gesture-support systems had the most benefit for the performance of distance collaborations.

Based on this motivation, Kirk and Fraser considered the two factors of gestures, format and output location, while constructing their remote gesture technologies. The combination of all these factors presents six different system configurations, as displayed in Table 1, which represent the conditions tested in their experiments. Figure 5 illustrates some of the gesture-support systems that they constructed, corresponding to some of the conditions in Table 1.

		Gesture Output Locations	
		Projected	***TV***
Gesture formats	***Hands only***	Projected hands, Figure 5 (a)	TV hands, Figure 5 (b)
	Hands and sketches	Projected hands & sketches, Figure 5 (c)	TV hands & sketches
	Sketches only	Projected sketches only	TV sketches only, Figure 5 (d)

Table 1: Comparison of possible gesture locations and formats [16]



(a) Projected hands only | **(b) TV hands only**

(c) Projected hands & sketches | **(d) TV sketches only**

Figure 5: Examples of David S. Kirk and Danaë Stanton Fraser's Technology Constructions to support collaborative physical tasks under different conditions [16]. Reprinted Courtesy of the Association ACM for Non-Commercial Machinery.

Kirk and Fraser's study was a series of three experiments, which were categorized by the gesture formats: hands-only, hands-and-sketches, and sketches-only. Each series of experiments included two trails, which were categorized by gesture output locations: projected hands and TV hands. Each of the three experiments was conducted using the same design, with the only difference between them being the output locations of remote gesturing used in each experiment. They kept the experiments consistently controlled, with a meta-analysis enabled that allowed the comparison of the two gesture location methods (projection and TV) to be carried out over all three of the sub-studies. A total of forty-eight pairs of participants collaboratively performed a Lego-assembly task under all these six conditions.

The evaluation resources used in Kirk and Fraser's study were: 1) Stage of completion—the progress made with the model after ten minutes (as measured in stages of the Lego kit completed); 2) The accuracy of the work achieved; 3) The post-test questionnaire response, which was given after each condition to assess a variety of inter-personal perceptions and opinions about task performance, including participants self-reported ratings of task difficulty, communication ease, personal productivity during the task, and understanding of partner's communications; and 4) Gesture location preference question, which was asked at the end of their two trials, so as to ascertain which gesture location method (projected or TV) the participants had preferred using. From the statistical analysis, the following results were found:

- **Completion Stages**: No significant difference was found between the two gesture location conditions; however, a significant difference was found between the hands-only conditions and the hands-and-sketch condition. They suggested that the conditions where a gesture was performed with a pen (either in isolation or conjunction to a representation of the hand) appeared to have lower levels of model completion, on average.
- **Accuracies**: No significant difference was found between all the conditions, meaning that very little perceptible difference was found between any of the conditions, in terms of accuracy of model assembly.

- **Post-Test Questionnaire**: No significant difference was found between the conditions. Thus, Kirk and Fraser suggested that varying the format of remote gesturing or the location at which it is presented has no discernable effect on tested task performance.
- **Gesture Location Preference**: Helpers showed very little preference for one method of gesture location over any other; however, workers tended to prefer projected hands over TV hands.

From the results above, a clear performance advantage is suggested based on the format of gesture representation was used to convey the remote gestures. Trend patterns clearly indicated that hand-based gesturing leads to quicker task performance than either alternative form of pen-based gesturing. A greater number of assembly tasks were completed after ten minutes under the hands-only condition than under either of the other gesture format conditions, with no loss in accuracy.

To make a conclusion, Kirk's project indicated that, when using the hand-based gestures (referred to as "the hands-only condition") in their remote-gesture systems, participants finished their tasks faster than using any format of pen-based gesture (referred to as "the hands & sketches condition" and the "sketches-only condition"). In addition, the hands-only condition had significantly better performance than the hands-and-sketches condition. Those results indicated the superiority of hand-based gestures in remote collaboration systems, when compared with pen-based gestures. This result is consistent with my study's result that the Gesture-Voice condition owns the best performance on average.

3.2 Remote Gesture Technology to Support Mixed-Presence Collaboration

Another broad category of remote gesture systems is Mixed Presence Groupware (MPG) [29], which is a software system connecting both collocated and distributed collaborators together in a shared visual workspace. The study of this new genre indicates that people focus their collaborative energy on collocated partners at the expense of remote partners; remote partners seem to cause an imbalance to collaboration. This problem is called *presence disparity*, and it is caused by the imbalance of visual, auditing, and other factors exuded by virtual embodiments rather than actual people. Mixed-presence collaboration has extended the arrangement to support multiple-party interactions. These systems exploit video projection techniques to support collaboration around the construction of shared representations (such as drawings) within collaborative design activities.

3.2.1 VideoArms Mixed Presence Groupware (MPG)

Anthony Tang and Saul Greenberg's Mixed Presence Groupware (MPG) scenario is described as follows [29]: "You lead a team of designers based in Seattle, and have scheduled a joint brainstorming session with another group in your New York office. This is possible because your company has special meeting rooms in each city's location, connected by audio and containing linked virtual whiteboards. This software allows one or more members to collaborate. So, either team can simultaneously draw ideas on the wall using the interface, where colleagues in either location can see those drawings as they are being created in real time." MPG usually represents collaborators within the workspace as entities by means of some type of embodiment—a virtual presentation of their bodies. In practice, MPG systems were built by connecting several distributed displays, each with multiple input devices, thereby connecting both collocated and distributed collaborators. The following figure shows a stylized example of an MPG system where three groups of collocated collaborators (top) work together in a shared virtual space (bottom).

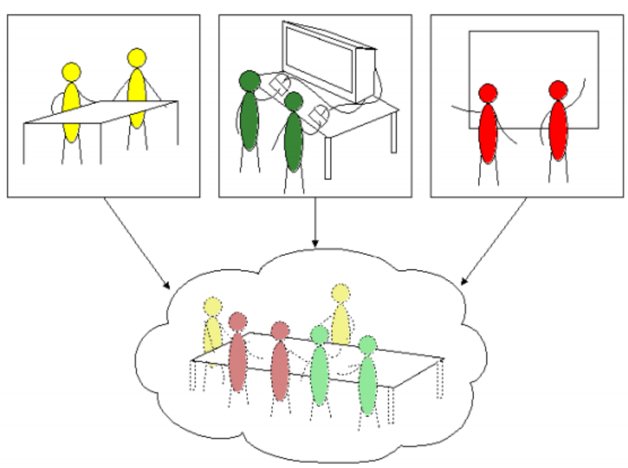


Figure 6: Three teams working in MPG over three connected displays (top), stylized as a virtual table (bottom) [29]. Reprinted Courtesy of Anthony Tang and Saul Greenberg for Non-Commercial Machinery.

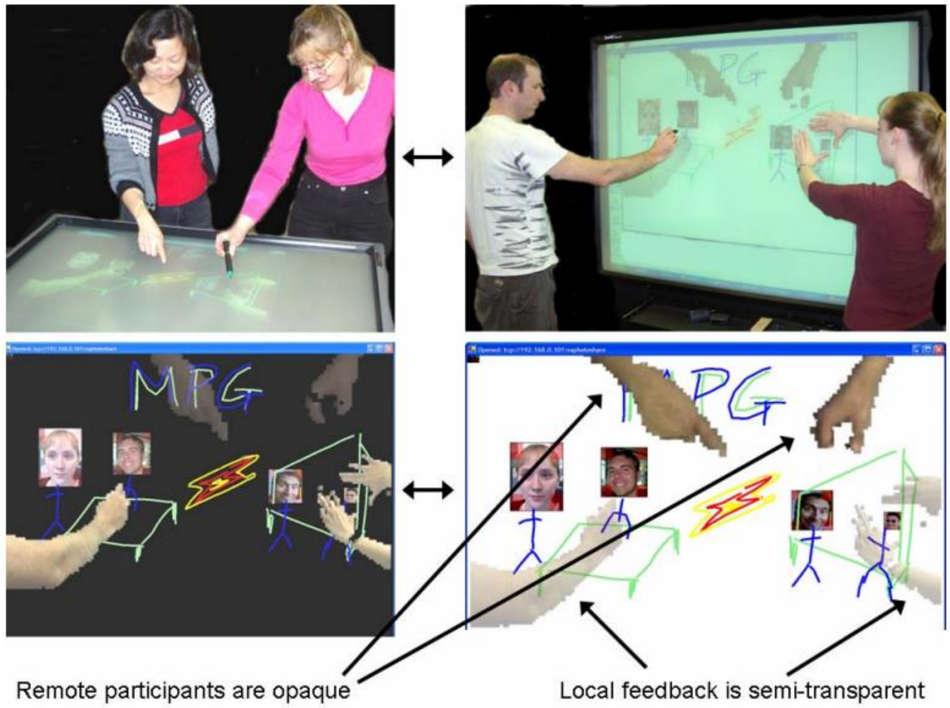


Figure 7: VideoArms in action showing two groups of two people working over two connected MPG displays (top) and a Screen shot of what each side sees (bottom) [29]. Reprinted Courtesy of Anthony Tang and Saul Greenberg for Non-Commercial Machinery.

In Tang and Greenberg's project, "VideoArms" [29] is an example of video-based MPG environments that digitally capture collaborators' arms as they work within the workspace using a camera, with the arms redrawn at the remote location. Figure 7 illustrates a sample session of VideoArms [29]. The top image of Figure 7 shows two connected groups of collaborators. (Each group works over a large touch sensitive surface—to the left is a front-projected touch sensitive horizontal DViT, while to the right is a rear-projected vertical SmartBoard. Each surface displays the same custom MPG application that will allow people to sketch and manipulate in stages, while displaying video embodiments.) The bottom image of Figure 7 also illustrates what users can see when using the VideoArms embodiment in this application. The design of this VideoArms technique aims to mitigate the problem of presence disparity in MPG. VideoArms digitally captures people's arms as they work over large work surfaces, and displays them as digital overlaps on remote displays. While doing so, VideoArms provides a rich means for collaborators to maintain workspace awareness of remote participants in MPG systems. One predominant feature of these systems is that they are based on the concept of shared interactive behaviors, where all parties share equally in the task and are working in similar roles, usually in some shared-presence disparity group collaborative design activity, or we may call this kind of activities group "collaboration design activities."

Tang and Greenberg did not report their evaluation in the published paper [29]. However, they stated that they had completed a preliminary study that demonstrated that VideoArms supports rich gestures and consequential communication across the link, thereby reducing presence disparity. The details of this preliminary study are not quite clear, however.

3.2.2 Mixed-presence tabletop interfaces

Another similar system is called Mixed-presence tabletop interfaces, which is also aimed to support collaboration between remote groups. Figure 8 [31] is an example of Peter Robinson and Philip Tuddenham's mixed presence collaboration [11] over digital artifacts using distributed tabletops. In this application, participants interact simultaneously to move and reorient the words to create poetry. They are jointly responsible for completing a single task. Visible in the photos are arm shadows, personal territories, and artifacts at arbitrary orientations. The benefits of tabletop interaction for mixed-presence and remote collaboration is discussed in this research. Tuddenham and Robinson stated that native use of embodiments, such as telepointers, leads to a disparity in the conversation dynamic whereby users are much more likely to interact with their co-located collaborators than with their remote collaborators, with a negative effect on collaboration. Tans et al. [11, 29] observed the use of mixed-presence whiteboards [6] and showed that richer embodiments, such as arm shadows, mitigate this effect. They review the roles played by physical bodies in collaboration and suggest that remote embodiments for mixed-presence collaboration should: 1) be controlled by direct input mechanics and allow remote collaborators to interpret current actions and the actions that led up to them; 2) allow remote collaborators to interpret gestures by capturing and rendering fine-gained movements and pictures; 3) appear in the workspace in order to convey gestures as they relate to the workspace; and 4) be visible not only to remote collaborators but also provide local feedback so that we might infer how our actions are interpreted by remote collaborators. Based on those four points, Tuddenham and Robinson concluded that supporting mixed-presence collaboration without a conversation disparity would support rich arm-shadow embodiments that follow these design guidelines, rather than using impoverished telepointers. From this analysis, they wish to support the natural tabletop awareness mechanisms of territoriality, orientation, and consequential communication. Tuddenham and Robinson's design guidelines for such systems, as well as the currently used Distributed Tabletops, can be customized to investigate various mixed-presence tasks.

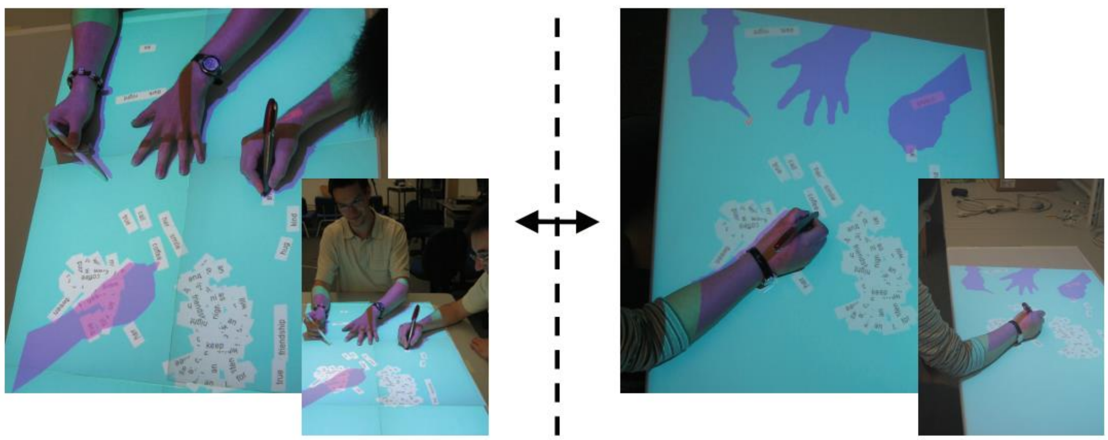


Figure 8: Mixed-presence tabletop collaboration over digital artifacts using Distributed Tabletops [31]. Reprinted Courtesy of the Association IEEE for Non-Commercial Machinery.

In the evaluation, Tuddenham and Robinson did not report controlled studies of Distributed Tabletops in the published paper [31]. However, they provided a formative pilot evaluation of the system using this application in two sessions with six participants in total. The evaluation of this system is based on the interviews with the users and the researcher's observations. In each session, participants reported that they felt responsive while using the system. They stated that, in the early observations, the system in use is promising.

To conclude, those two projects in support of Mixed-Presence Collaboration did not have detailed evaluation processes, while Fussell, Kirk, Drew, Cornelius, and my project did have detailed evaluation processes in place. Although there was some evidence that hand shadows or videos would improve presence and collaboration, such evidence was not qualified in a controlled study to verify if hand videos really helped virtual collaboration.

3.3 Foundations of the Current Work

This subsection describes two projects on which the current project is directly based: 1) *Virtual Collaboration in Conceptual Design*, led by C J Cornelius [37]; and 2) *Virtual Collaboration in Traffic Accident and Disaster Management*, led by Daniel Drew [10]. These two studies were carried out in the same lab using the same equipment and some of the same software; several new interfaces and improvements were incorporated from Cornelius' original software for Drew's study. The project described in this thesis, "Hand images in virtual spatial collaboration for map-based planning activities”, is a follow-up to those two projects. I led my thesis project jointly with Andrew Ahlfield. The current work uses the same physical equipment. Many of the questions explored in this thesis arose from unanswered questions identified in Drew’s project. This subsection will describe the goals, system setups, experimental designs, evaluations, results, and conclusions from these two studies.

3.3.1 Virtual Collaboration in Conceptual Design

The project, *Virtual Collaboration in Conceptual Design*, was led by C. J. Cornelius. The goal of this project was to answer whether hand images are necessary at all or if a simple cursor would be sufficient for a collaborative drawing, in which two users in multiple locations can connect and draw using a simple drawing tool. The two users are equal in task scenarios, rather than being “helper” and “worker.” The underlying assumption in Cornelius’ work is that hand gestures, used in the context of shared sketches, drawings, and other spatial representations, are an important form of spatial communication in engineering design and other spatial tasks, referred to as "Conceptual Design" by Cornelius.

Her physical setup is shown in the left side picture of Figure 9, including a projector, a camera, an infrared-sensor, an infrared pen, and a drawing surface on each side of the wall. This physical setup is based on the scenarios of distance collaborations for product design tasks. Her software interface allows multiple participants to engage in joint spatial drawing tasks even when the participants are physically separated; they each can hear the other, each can manipulate a joint drawing, and each can see the others' hands and gestures overlaid on the drawing surface. Two tools were created to make up the software interface: one for virtual sketching (shown in Figure 10) and the other for virtual sketching that enables the sharing of hand gestures. Thus the software on each side of the wall has three parts, as pictured in Figure 9, to the right: 1) The shared drawing surface on which the participants can use an infrared pencil to draw (the infrared sensor which is used to react to their infrared pen is located on the ceiling, next to the projector); 2) The remote virtual participant's shadow is captured by the camera and then projected on the local participant's virtual workspace to improve the sense of presence; 3) The physically present participant.

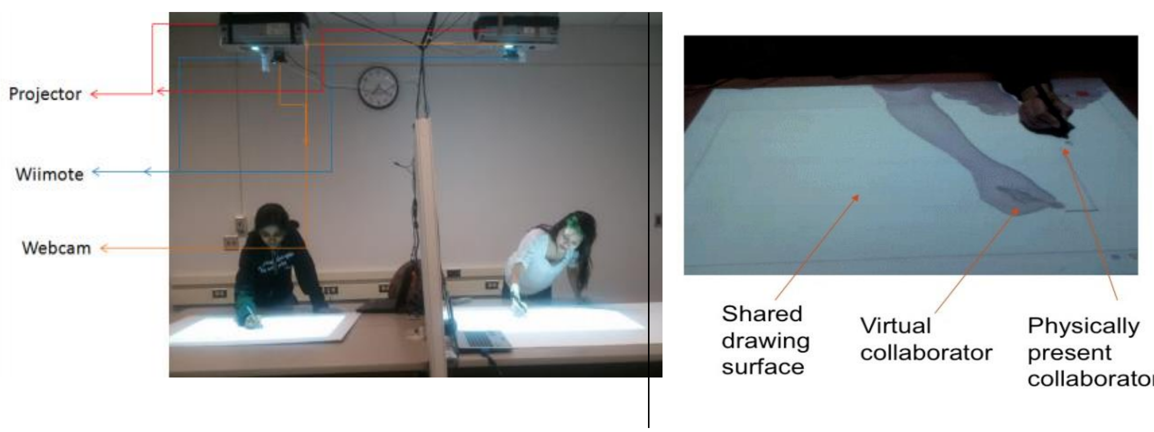


Figure 9: Physical Hardware Set-Up (the left picture represents the overall setup; the right picture is the enlarged view of the interface surface, which shows the virtual sketching tool) [37], courtesy of C. J. Cornelius.

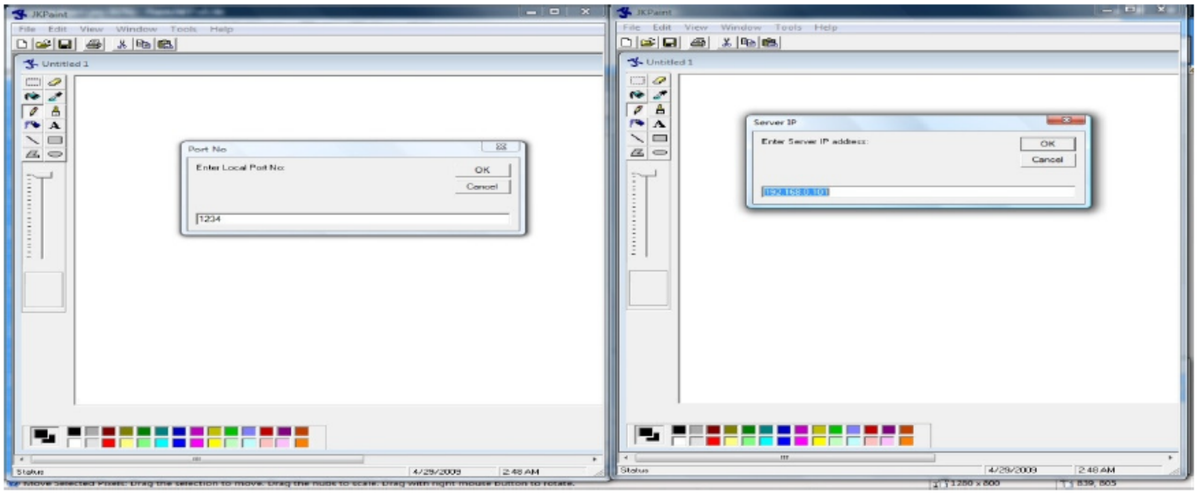


Figure 10: Virtual Sketching tool [37], courtesy of C. J. Cornelius.

Her study was conducted on twelve pairs of subjects (twenty-four subjects total). They were asked to jointly complete three engineering design tasks, using three different collaboration interfaces. The order of tasks, interfaces, and presentations was systematically varied. Each of the twelve pairs performed three tasks using three different approaches for collaboration:

- **Face-to-face**: Both participants are placed in a single collaboration workspace, sharing one infrared pen for drawing;
- **Virtual sketching only:** Participants each sit in separate collaboration cells, where they can use the screen-sharing portion of the tool to create a joint drawing and make gestures of a limited nature by moving the cursor;
- **Virtual sketching with hand images**: Same as interface (b), except that participants can also see each other's hands and gestures (or any other objects they choose to place in the camera's view).

The evaluation of the software was based on comparing the collaborations for a spatial design drawing created when users collaborated by using three different methods: face-to-face, virtual sketching, and virtual sketching with the collaborators' hand gestures projected on the drawing surface. The data collected were: 1) Percentage of Time Spent On-Task, which were used to determine how much of each users' time was spent in on-task versus off-task activities (she used the percentage of time spent on-task as a measure of how much the tool may have interfered with the users' ability to get work done); 2) Task-completion time; and 3) Questionnaire data, which includes NASA's TLX Workload Measurement (to measure the mental demand, physical demand, temporal demand, performance, and frustration level), and questions about preferences. From the statistical analysis, Cornelius found the following results:

- **Percentage of Time Spent On-Task:** Subjects spent a significantly larger proportion of their time on-task when using virtual collaboration with hand images than in either of the other two conditions.
- **Task-completion time:** On average, participants completed the tasks fastest under the face-to-face condition (mean = 5.355 minutes); second fastest, under the virtual sketching with hand-images condition (mean = 7.380 minutes); and slowest, under the virtual-sketching-only condition (mean = 8.755 minutes). Cornelius did not report the ANOVA analysis of task-completion time.
- **Questionnaire**
    1) **NASA's TLX Workload Measurement**: There was no significant difference found between face-to-face collaboration and virtual sketching with hand images in the 6 TLX categories, but only the virtual sketching was significantly more demanding on all measures;
    2) **Preferences Questionnaire**: Most participants preferred face-to-face or virtual sketching with hand images.

The above results indicated that users' performance in the face-to-face collaboration was almost identical to virtual collaboration with hand images except for the task-completion time. Virtual collaboration without hand images was significantly more difficult. Based on those results, she concluded that: 1) rich hand gestures conveyed with the hands in face-to-face settings are important in joint spatial tasks; 2) the addition of hand images projected on a joint work surface can restore effectiveness to levels similar to that of face-to-face collaboration; and 3) a robust setup of the software enabling transfer of rich hand gestures could be used for a virtual collaboration that gets a distant team member to be present in a shared virtual workspace.

To conclude, Cornelius' results are consistent with my study's results, which show that when gestures were added to the virtual-map interface, the remote conditions became statistically distinguishable from the Face-to-Face condition. Furthermore, my study indicated that, the Gesture-Voice condition was slightly superior to Face-to-Face, in terms of task-completion time, mental demand, frustration level, and the seamlessness of collaboration.

3.3.2 Virtual Collaboration in Traffic Accident and Disaster Management

The project, "Virtual Collaboration in Traffic Disaster Management," was led by Daniel Drew. This project focused on the modes of communication that support collaboration between distant collaborators managing traffic incidents or disasters. The goal of his work was to develop demonstration technology that can overlay hand videos on spatial images, such as traffic maps, and assess the impact of this technology on virtual collaboration. His work has explored the degree to which gestures impact collaborative effectiveness in the management of traffic incidents, with the goal of informing the design of tools to support virtual collaboration in this domain.

This project is using the same physical setup as Cornelius' project [37]. However, the software system is different. Under the domain of traffic incidents and disaster management, much of the information needed to communicate refers to spatial locations on a map. Therefore, I developed a virtual-map interface and a marking interface to facilitate the distance that collaborators need for communication in traffic accident and disaster management, as shown in Figure 11. The whole interface combines together the Google Map API and the marking tools. This interface allows participants at two different locations to view the same map, jointly marking the map, and seeing each other's markings. Also, to aid in Drew's software development, I improved the algorithm that captures hand images from Cornelius' project [37]. To make things easier for users, I reduced the hand-image noise that existed in the old software, as shown in Figure 12. Because of my improvements, the captured hand images have no shadow at all and the hand video is much smoother than before, as shown in Figure 13.

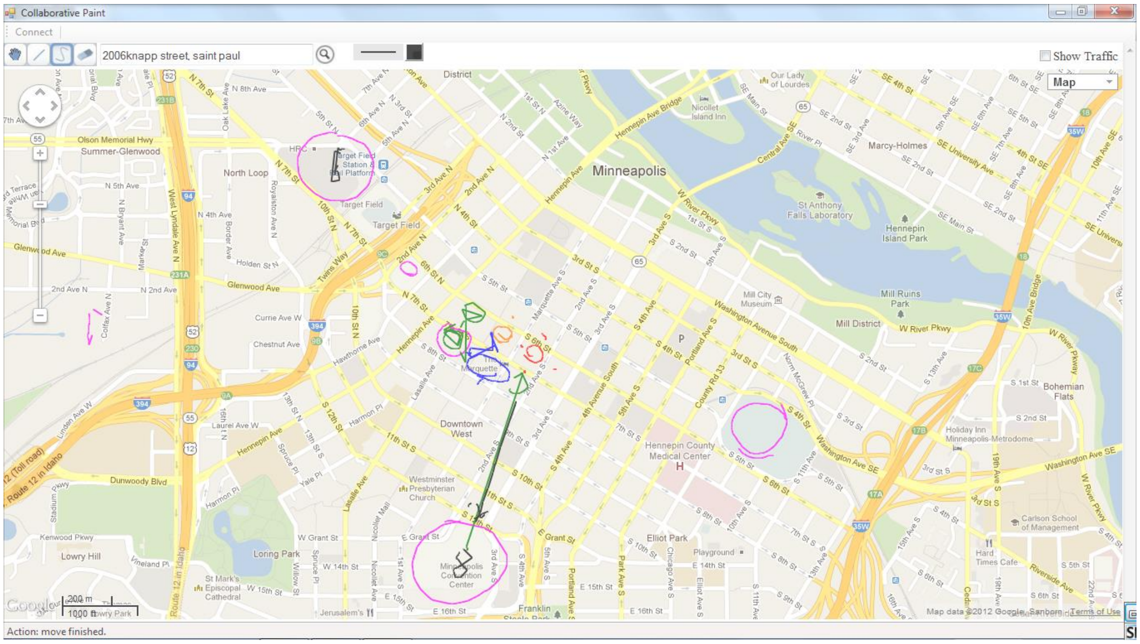


Figure 11: A combination of the virtual-map interface and the marking interface.

Figure 12: A hand image transferred by the old hand-image substraction software in the Virtual Collaboration in Conceptual Design project (pictures are captured while the hands are moving) [37]

Figure 13: The hand image transferred by my new hand-image substraction software in the Virtual Collaboration in Map-Based Planning project and Traffic Incident & Disaster Management project (pictures are captured while the hands are moving; the pen shadow is the infrared pen used in the collaboration.)

Drew's study was conducted with eighteen participants who worked in pairs to solve three traffic incident scenarios using three different interaction approaches:

- **Face-to-face**: Participants worked together by marking up an electronic map projected on the table in front of them;
- **Virtual sketching only**: Participants were separated by a soft wall while they worked together on the electronic map with electronic drawing tools;
- **Virtual Sketching with hand images**: Same as 2, with the addition of the partner's hand images projected on the map.

Participants were video recorded. The questionnaires were given to participants after each trial to evaluate workload, positive interactions, team behaviors, connection to teammates, and frustration-level.

The evaluation of the software was based on comparing the collaborations for traffic incidents and disaster management tasks when users collaborated by using the above three different approaches. The data collected were: 1) the time to complete the tasks; 2) the questionnaire: a) NASA TLX Workload Measure, with b) questions about positive interaction (including three questions on encouragement, the received positive feedback, and the positive feedback given), c) questions about team behavior (measured by four questions on discussion quality, joint planning, team perception, and joint work), and d) questions about connection to teammates (using three questions to assess presence, social distance, and engagement); and 3) video codings: the breakdown of time spent "inside" versus "outside" the shared electronic interface and eight categories of actions (Gaze-only, Manipulate, Speak, Mark, Gesture, Gesture & Speak, Mark & Speak, and Manipulate & Speak) for each interface. Through the statistical analysis, Drew found the following results:

- **Time to Complete the Tasks:** Participants completed the tasks significantly faster when working face-to-face than when using hand images.
- **Questionnaire**
    1) **NASA TLX Workload Measure**: A significant main effect on the interface. Additionally, the pairwise comparisons showed a significant difference on NASA-TLX

measures of temporal demand; participants perceived significantly more time pressure when using hand images than when working face-to-face.

2) **Positive Interaction:** No significant difference between interfaces.

3) **Team Behaviors**: A significant main effect on the interface; participants perceived significantly more teamwork when using the hand-images interface than when using the separated interface or the face-to-face interface. However, the pairwise comparisons did not show any significant differences among individual components of this aggregate measure: discussion quality, joint planning, team perception, and joint work.

4) **Connection to Teammate:** A significant main effect on the interface. Additionally, the pairwise comparisons showed that participants felt significantly more present and less disconnected when working face-to-face than when using the hand-images interface or the separate interface.

- **Video Codings**

  1) **"Inside" versus "Outside":** No significant differences between the interfaces. However, Drew assumed that participants seated on opposite sides of the wall would rely more heavily on the shared interface, since they had no other means to communicate visually. This assumption contradicted the actual result. He suggested that if the participants truly were not frequently using the shared electronic interface, this might explain why no significant differences in the NASA TLX measures appeared for these two conditions. Therefore, he raised the question: "Were Participants Actually Using the Shared Electronic Interface?"

  2) **Eight categories of actions:** Under the face-to-face condition, participants spent significantly less time speaking than in the other conditions and significantly more time gesturing (while speaking) and less time marking than in the separate condition. There are no significant differences between face-to-face and hand images in the total time spent gesturing (while speaking). However, further details about the hand-image

condition need to be explored. Therefore, he raised the question: "How was Time Spent?"

The result of Drew's project indicated that: 1) The ability to gesture to one's virtual teammate, by pointing or making motions over a shared electronic map, may improve team behaviors in the domain of management of traffic incidents and disasters; 2) The presence or absence of gestures in the interface changed the way in which people communicate: when participants interacted face-to-face they spent much time gesturing; when they could no longer see each other in a virtual interaction, they compensated by talking and drawing more; when gestures were also taken away, they gestured very little (relative to face-to-face interactions).

To conclude, Drew's results did not show a clear advantage to adding hand videos over a virtual drawing tool for the task of traffic incident management. In particular, Drew did not see a significant improvement in performance like Cornelius did when hand images were added to the virtual interface. From the analysis of Drew's experimental design and the experiments' video records, we considered three possible reasons why Drew's results did not show a clear advantage to adding hand videos, and raised three questions to be explored in the current project. The possible reasons and resulting questions are:

1) *Participants were not actually using the interface all the time.* From the observations made during Drew's experiment, we found that the participants were not actually using the shared virtual-map interface all the time; in contrast, they used the paper maps with the task descriptions a lot. Also, participants were more likely to finish each other's work separately and then combine their results all together, instead of collaborating on the shared interface. We supposed that this might have been because the traffic incident and disaster management tasks need broader expert knowledge than do the map interactions. Another possible reason is that the participants were given paper maps, which people rarely use anymore in everyday life. Based on the above analysis, we raised the question: "**Will the various virtual collaboration tools (markings, hand images) help more,**

if we design the interface so that participants can only collaborate through the shared virtual interface?"

2) *The addition of the sketches function slowed down the performance of the hand-image interface.* Drew's project showed that the addition of hand image was not significantly better than the sketches-only condition. Kirk and Fraser's project for physical assembly tasks [16] also showed similar results, that no significant difference existed between the two formats of sketches, which involved conditions (the hand-images & sketches condition and the sketches-only condition). However, unlike Drew's work, Kirk and Fraser's experiments included a condition with hand-images only. Kirk found that the hand-images-only condition owned significantly faster task performance than either of the formats with sketches-involved conditions (the hand-images & sketches condition or the sketches-only condition). Therefore, Kirk and Fraser suggested that to add sketches actually slowed people down in virtual distance collaboration. Thus, we encountered the question, **"How will the performance be impacted when the hand images are provided, without the sketching (marking) functions?"**

3) *There is no comparison with Voice-only condition.* In Drew's experimental setting, there is no comparison to the traffic managers' existing methods of distance collaboration, a Voice-only condition (by phone, through which remote collaborations are usually carried out in everyday life). We were unable to see the difference when comparing the various supportive tools in basic conditions. However, those differences are meaningful, especially when the differences between various supportive conditions are insignificant. But, Drew's project did not have this basic, Voice-only condition, which could be another reason for his inadequate result. So we raised the question, **"What will the differences be when various supportive conditions (Mark-Voice, Gesture-Voice, Mark-Gesture-Voice) are compared with a Voice-only condition (e.g., by phone)?"**

To answer the above questions raised from Drew's project, the current project was developed as a new iteration of Drew's project, with some experimental changes and additional conditions, in order to better control the experiment and clarify the impact of hand images in virtual collaboration. The following three points are the major changes in the current project to address the raised questions from Drew's project:

1) To address the first question about the experimental design:

    a. The tasks were refined to pertain to map interactions and everyday life, which do not require traffic expert knowledge, but only a user familiar with Google Maps. The tasks were simplified, with no clear "right" answer.

    b. There is no paper map provided in our current project. Therefore, the participants will be more focused on the virtual-map interaction.

2) To address the second question about the impact of hand images on virtual collaboration, we added the Gesture-Voice condition into the current project.

3) To address the third question about the results of a comparison with a common virtual collaboration (e.g., by phone), we added the Voice-only condition into the current project.

4. System Development

In my current project, the physical set-up is the same as the conceptual design of the physical set-up of Cornelius' project [37] (shown in Figure 9). However, the software system of this application is different. The software is made up of three layers of interface. Those three layers are overlapping and the top two layers are 30% transparent:

- **A Virtual-map Layer**, referred to as the central layer in Figure 14. This layer was developed for the base requirement of map-based planning tasks and was used as a base function in the experiment. Using this interface, the collaborators can zoom in or out and drag this virtual-map with the traffic view enabled or disabled. Also, the virtual maps can be either synchronous or not, within the remote collaborations, according to the experimental conditions.
- **A Marking Layer**, referred to as the bottom layer in Figure 14. This layer was designed to improve the collaborators' cohesive communication and was used for the marking function in the experiment. This marking layer can interact with the virtual-map layer, synchronizing the existent markings with the virtual-map layer. Thus, the combined interface of the virtual-map layer and the marking layer looks like a manageable synchronized virtual map with marking functions, as shown in Figure 11. Also, the markings can be either synchronized or not, within the remote collaborations, according to the experimental conditions.
- **A Hand-images Layer**, referred to as the top layer in Figure 14. This layer was also designed to improve the collaborators' distance communication and was used for sharing gestures in remote collaborations. The improvement is based on the hand-image algorithm in Cornelius' project [37]. After my improvement, the hand images projected for this layer are much smoother and the noise is decreased, as shown in Figures 12 and 13.

Figure 14 on the next page illustrates the combination of the three layers.

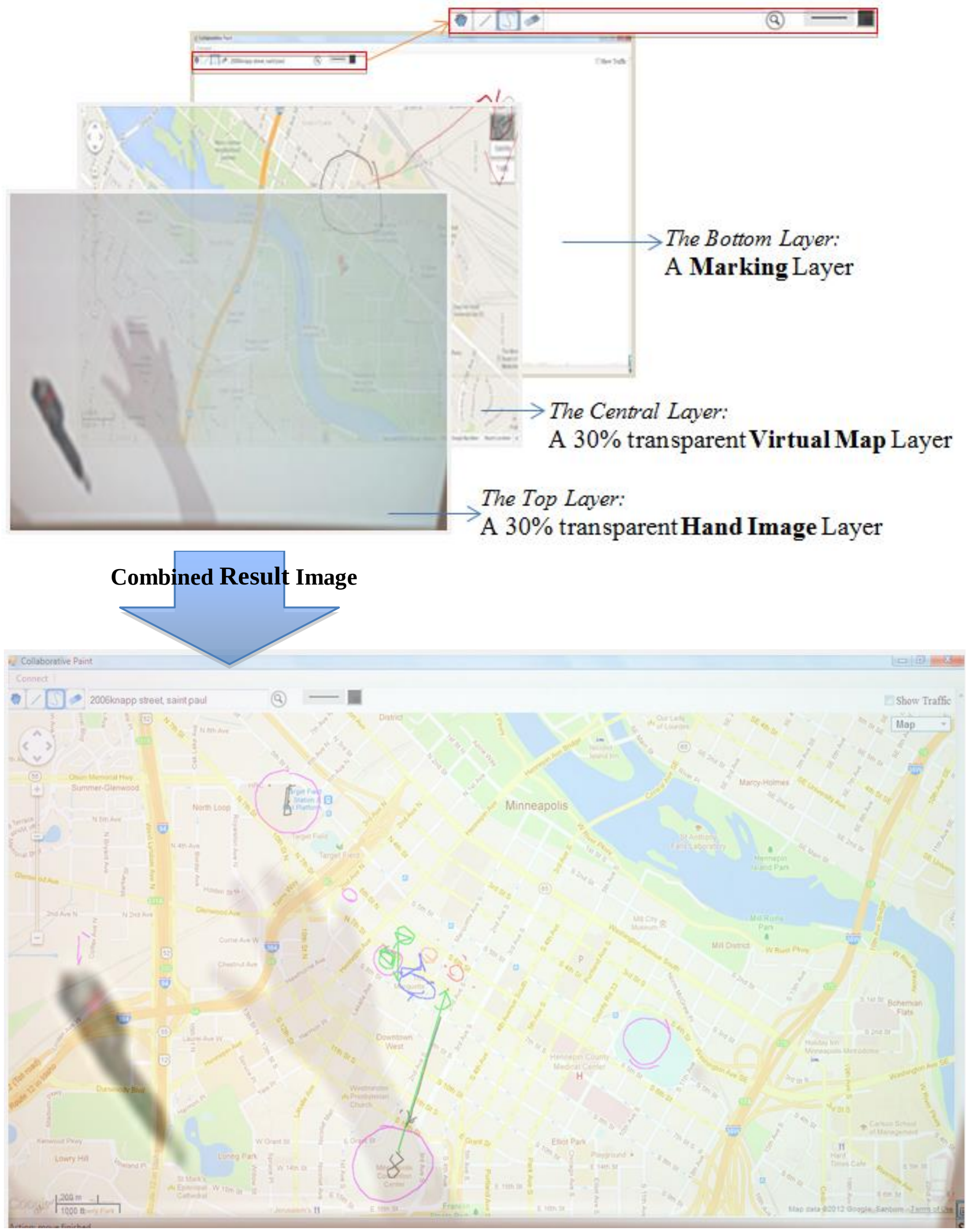


Figure 14: Three Layers Comprise the Virtual Workspace

On the other hand, the synchronous process of those three layers is described in the following sequence:

1) Detect the changes or filter the information of the hand images from the local site;

2) Transfer the information changes as coordinates and events or images onto the remote site;

3) Interpret the transferred information to strings and commands, and then replicate the local interface's changes or hand and arm images on the remote site.

Each site was running both as a local site and the remote site at the same time. The representation of this synchronous model with details is shown in the following figure.

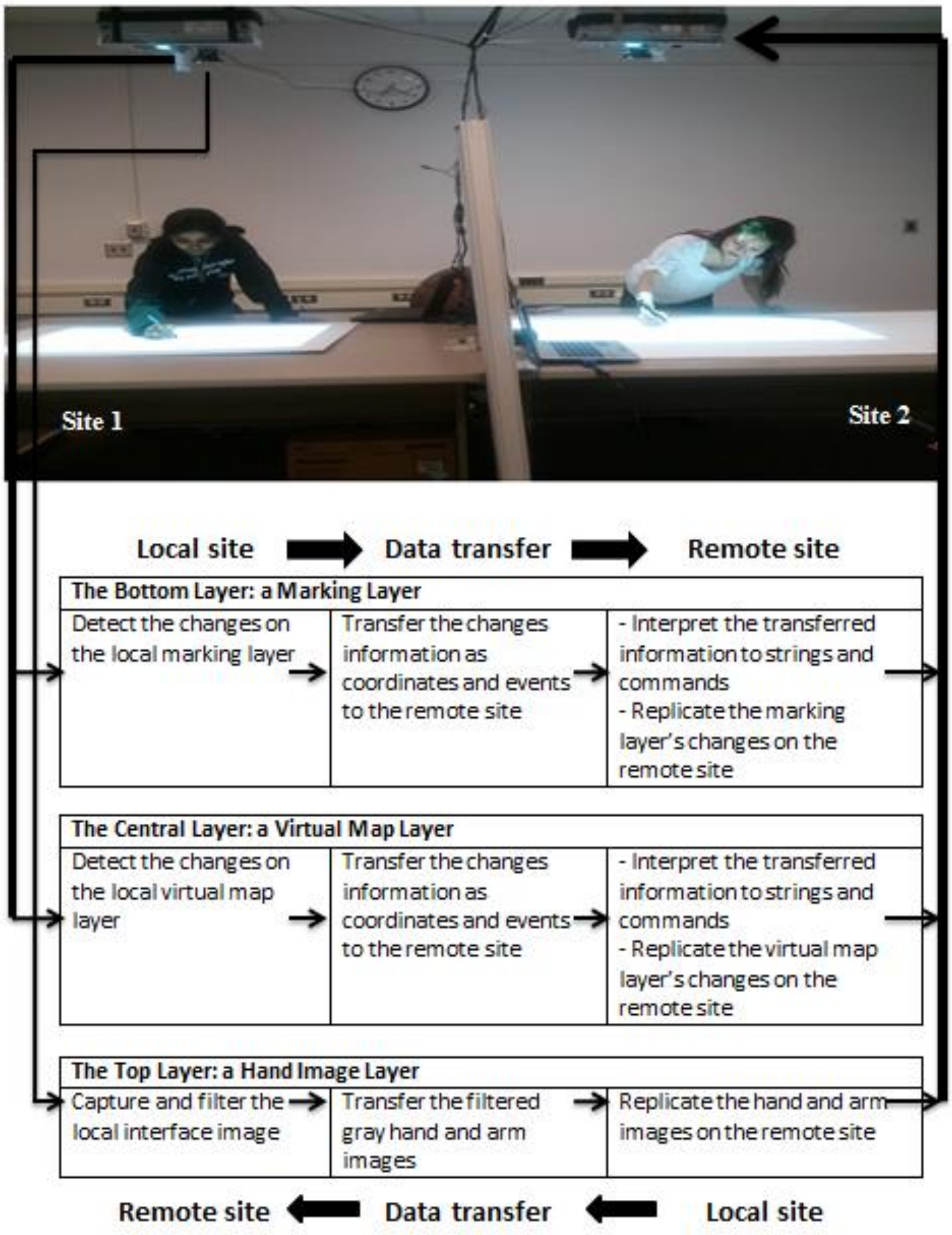


Figure 15: Synchronous Model Representation of the Design

4.1 Development of the Virtual-map Layer and the Marking Layer

For current map-based planning activities, distance collaborators usually only look at separate maps and discuss their ideas over the telephone. Since map information is complicated to describe, such collaboration could be quite time-consuming and frustrating. Thus, in adapting the sharing concept from *Google Document Share,* the first aim of the current project is to provide distance collaborators a shared virtual map, and enable them to share their maps similarly to the sharing method of a Google document. Using this synchronous virtual-map interface, the collaborators will look at the same map information simultaneously and feel like they are all using the same map. This interface integrated the following functions from Google Map *Application Programming Interface* (API): 1) **Drag Hand Button—**to change the map's viewpoint; 2) **Search Button—**to show the map of a certain address; 3) **Zoom and Move Controller—**to zoom in or out of the map and to move the map in all four directions, south, north, east, and west; 4) **"*Show Traffic*" Checkbox—**to show the traffic congestion on the map if checked; 5) **Map Dropdown Options—**to change the map aspect for terrain or satellite view.

In addition, to helping distance collaborators express their map-based information faster and to save time explaining map complications only using words, the next step will be to build a technological feature that can allow collaborators to mark on the virtual map and to share their sketches and markings, and then, to test the performance of this marking technology through our experiment. Based on this goal, we developed a marking interface that can be placed under the 30% transparent virtual map, so that the collaborators might feel as if they were marking on the virtual map. Using this marking interface, distance collaborators can mark on the virtual map and explain their information or ideas at the same time; collaborators can then feel like they have opened a dynamic Google Map using simplified MS Paint software. This marking interface is synchronous with the virtual-map interface. So, when collaborators move or zoom in and out on the map, the location of any map marking stays the same. I integrated the following marking tools to the marking interface, mainly using JavaScript and HTML programming languages: 1) **Straight Line Button—**to mark a straight line; 2) **Circle Line Button—**to mark any shape; 3) **Erase Button—**to erase all the markings after confirming; 4) **Line-size Dropdown Options—**to choose a line size for marking; 5) **Color Dropdown Options—**to choose a color for marking.

From a programming perspective, the various functions of the virtual map and the marking interfaces are developed by two major programming-language technologies: HTML and JavaScript. The synchronous functions of those two interfaces are implemented by C# language and Socket programming, using a Client-Server Model. The following sub-sections will explain those programming technologies.

4.1.1 HTML

HTML was used to create the web page layout and display the information on the interface. HTML allows images and objects to be embedded and can be used to create interactive forms. It provides a method to create structured interface formats by denoting structural semantics for user-interactive function areas and virtual-map interface areas. Embedding HTML with JavaScript affects the behavior of HTML interfaces, by creating interactive content.

HTML documents are composed entirely of HTML markup elements. Structural markup language describes the purpose of the text content. Presentational markup describes the appearance of the text, regardless of the purpose. The attributes of an element are name-value pairs. For example, the ID attribute, as a structured markup element, provides a document-wide unique identifier for an element; the STYLE attribute, as a presentational markup element is used to attach a sub-textual explanation to an element. So, a STYLE attribute was used in the current system to display the images of all the interfaces; an ID attribute was used to identify the specific interfaces, such as the various buttons and map coordinates that interact with JavaScript functions, and implement the various functions of the map and markings and synchronize the interfaces. To conclude, HTML was used here to develop the structures of all the interfaces, including the Drag Hand Button, Search Button, Zoom-and-Move Controller, “Show Traffic” Checkbox, Map Dropdown Options, Straight Line Button, Circle Line Button, Erase-All Button, Line-size Dropdown

Options, Color Dropdown Options, and Map Canvas. Those buttons, controller, checkboxes, dropdowns, and map coordinates are in turn working with JavaScript functions to have various interactive effects on the whole map-based interface.

4.1.2 JavaScript

JavaScript is an interpreted computer programming language. It was implemented as part of the virtual-map interface and the marking interface. Firstly, I used this language to detect any action on the virtual-map layer and the marking layer and then updated those two layers accordingly. In detail, all the actions on the interfaces, such as “Click the IR pen” or “Hold and click the IR pen” are detected and recorded as strings using JavaScript programming. Then I programmed JavaScript languages to analyze those strings and create commands to implement various functions, such as marking or dragging. Using those created commands, I implemented the updates on the local site. Secondly, I used JavaScript programming to transfer all the recorded strings from the local site to the remote interface, and then replicated those commands on the remote site using the same logic. Through this process, I implemented the synchronization of the local interface and the remote interface.

There are several important JavaScript functions used in the programming process. The first important function is the *initialize()* function which sets the default condition of the shared virtual map. The other more than twenty different functions can be divided mainly into two kinds of uses: 1) Manipulating page elements (i.e., animating, resizing, and moving elements, etc.); 2) Transmitting information about the user’s behaviors. Criteria functions of handling were developed to implement these usages. The most important criteria functions here are the *draw()*function and the *Junk()* function. The *draw()* function records the latitude of a line’s length and location, then marks them on the marking interface to make this virtual map personal and manageable. The *Junk()*function works together with the *draw()* function, making the record have different line colors and line weights. The most complex functions are color dropdown options and line-size dropdown options. Those two functions were developed to implement the options of presenting different unique marks whenever the user is marking.

4.1.3 Client-Server Model and Socket Programming

This project's collaboration system is bidirectional. The directionality was implemented using a symmetric Client-Server Model. In a Client-Server Model, to establish a communication channel between two processes, one process takes the initiative, while the other waits; the process initiating the communication is the client, and the process waiting for the communication to be initiated is the server. In this project's peer-to-peer virtual-map collaborative communication, each site of the collaboration system has acted both as a client and as a server, so we call it a symmetric Client-Server Model. The client and server have worked together, to form my symmetric, distributed, collaboration system.

The endpoint in this symmetric distributed system was implemented by a regular socket or a network socket for disambiguation. The data transmission between two sockets was organized by communications protocols—TCP/IP—and implemented in the operating systems of the participating computers. Application programs wrote to these sockets and read from these sockets.

4.2 Improvement of the Hand-image Layer

In Cornelius' project, her hand-image software projected hand and arm images, but was very noisy when the hands were moving; also, the hand videos appeared jerky and lagged behind the actual motion. Figure 12 illustrates a series of screenshots showing the noise that Cornelius' hand imaging software projected. If this software was implemented under a simple background interface like a whiteboard interface, this may not have been a serious problem. But under our complex virtual-map background (shown in Figure 11), this software may confuse and frustrate the users with the many details already present on the map and with the markings. Therefore, the hand-image interface of the old program must be optimized.

Through analysis, the hand image noises that Cornelius' software projected are history images of the hands. The fundamental reason behind those noises and the hand videos' choppiness is that her hand-image software transferred whole interface images without any filtering, optimization, or compression. The program needs to transfer large amounts of information and therefore, the frame frequency needed to be reduced. (Frame frequency is the frequency at which the video camera produces unique consecutive images called frames. The human eye and its brain interface, the human visual system, can process ten to twelve separate images per second, perceiving them individually. A video that produces more than twelve frames per second is called a real-time video.) Apparently, the frame frequency of Cornelius' software was not in real time. History images were captured and reflected as hand-image noise and the videos were jerky. To delete those noises and improve the smoothness of hand-image video, the software program should transfer only the hand and arm images, instead of whole images. Through changing the type of images transferred, we can release the load that needs to be transferred, and can increase the frequency of image-frame capturing and transferring, enabling the software to project real-time hand videos without shadowed hand images. Based on the above analysis, new software was developed based on the *Image Subtraction and Overlaying* algorithm. This algorithm subtracts the hand image from the background image and overlaps the subtracted images onto the remote site [40].

The development of the hand-image subtraction algorithm arose through the study of human skin color [39]. Previously, studies in hand tracking have approached recognition of skin-color distributions that have been undergone within a certain set of conditions. Adapting the concepts from those previous approaches, and in order to recognize selected hand objects from the shadows, I compressed and captured image information by transforming the colorful captured image into a recognizable grey image and then analyzed the grey degree of the whole image. Based on those analyses, I developed an optimization algorithm to recognize the hand and arm objects through setting a suitable grey filter value and adjusting the brightness of the background images. Thus, the subtracted hand images would look like a pure grey shadow, which makes the peer hand image even clearer and easier to recognize. Also, since grey images are light-loaded information, so, this process will release a lot of the loads to the transferring process and thus, the frame frequency will also be increased. To summarize, after completing the above processes, the new hand-image software can: 1) Subtract the hand images from all the other images on the combined or overlapping interfaces; 2) Transfer only grey hand-image information; and 3) Project real-time hand videos with no noise. Figure 13 illustrates a series of projected hand and arm images with no noise output during the peer hand movement.

5. Experimental Design

After the development, the next problem was how to test this system. I participated in the development of the experimental design with Andrew Ahlfield to explore the hypothesis question of this project, which is, "By providing virtual collaborators with a more natural vehicle, hand images, could performance improve , workload decrease, and team cohesion increase?" In our experiments, participants were recruited to complete five tasks in pairs, under the variable of five different conditions: 1) **Face-to-Face**; 2) **Gesture-Voice**; 3) **Mark-Gesture-Voice**; 4) **Mark-Voice**; and 5) **Voice-only**. The tasks are all map-based planning tasks. Participants had complementary task information to collaboratively complete the tasks. The order of these tasks and conditions changed systematically to avoid ordering effects in the data. Then we statistically analyzed the task-completion time to measure the performance and the questionnaire results to measure participants' workload and team cohesion.

5.1 Participants

Twenty-eight participants were recruited to join in the experiment. They were drawn from undergraduate or graduate students who are familiar with Google Maps. They worked in pairs to solve five different tasks using five different interface conditions. During the tasks, they were asked to use the tool to interact with each other from different sides of the wall, and provide feedback about its weaknesses and strengths through the improved questionnaire and the interviews.

Before the actual test, eight participants were asked to run four pairs of beta tests, which were used to help the experimenters avoid mistakes, to identify the vagueness in the task descriptions and the questionnaire, and to become familiar with the experimental process and the task-completion time. The participants from the beta test were asked to provide feedback about the whole experiment, the task descriptions, the questionnaire, and possible improvements. Feedback from this step will be used to improve the experimental design prior to developing larger studies. After the beta testing, twenty participants were asked to run ten pairs of actual tests.

5.2 Experimental conditions

We created five experimental conditions, using five different versions of the virtual-workspace tools. Figure 16 shows the scenario of the Face-to-Face condition, which was used in local collaborations to make a contrast to the remote collaborative conditions. Figure 17 shows the scenario of the remote collaborative condition, which was used in all the other conditions.

Figure 16: The Face-to-Face condition

Figure 17: The remote collaborative condition

Those five conditions are:

- **Face-to-Face** (in contrast to remote collaborative conditions): Two participants sit on the same side of the wall and share the same virtual-map and marking interface, using one workspace. They solve the problem by discussing things with each other and by manipulating (drag and zoom in and out) and marking on the virtual map using line and circle tools on the face (shown in Figure 16). Condition 1 uses both a virtual-map interface and a marking interface.
- **Voice-only** (the base condition of remote collaborations): Two participants sit on opposite sides of a wall using separated workspaces. They can only use their voices to communicate with each other. They can also manipulate (drag and zoom, in and out) and mark on their own virtual maps. Condition 2 used two virtual-map interfaces and two marking interfaces, all of which are unshared.
- **Mark-Voice**: Two participants sit on opposite sides of a wall using separated workspaces. They can use their voices to communicate with each other but perceive each other only as cursors on the shared virtual map. They can also manipulate (drag and zoom, in and out) the virtual maps and mark on the virtual maps using line or circle tools. Condition 3 used two shared virtual-map interfaces and two shared marking interfaces.
- **Gesture-Voice**: Two participants sit on opposite sides of a wall using separated workspaces. They can use their voices to communicate with each other and perceive each other by the hand and arm images of their partner on the shared virtual map. They can also manipulate the virtual maps (drag and zoom, in and out), but cannot mark on the virtual maps. Condition 4 used two shared virtual-map interfaces and two hand-image interfaces.
- **Mark-Gesture-Voice** (a combination of Mark-Voice and Gesture-Voice conditions): Two participants sit on opposite sides of a wall using separated workspaces. They can use their voices to communicate with each other and perceive each other by the hand and arm images of their partner on the virtual map. They can also manipulate (drag and zoom, in and out) and mark on the virtual maps using line or circle tools. Condition 5 used two shared virtual-map interfaces, two shared-marking interfaces, and two hand-image interfaces.

5.3 Tasks Design

Five map-based planning tasks were designed for the experiment. They are interesting and close to life. Also, most of them were in Minnesota. We focused on: 1) making it possible for participants who are familiar with Google Maps to complete the scenarios in less than fifteen minutes; and 2) structuring to require interaction through the interface. Also, for the purpose of collaboration, each task has two versions of the descriptions, which contain separate, but complementary information, for Participant I and Participant II. The participants need to collaborate in order to complete the task. The five tasks' titles and goals are listed below:

- **Bomb Threat/Evacuation**: The IDS Center in downtown Minneapolis is subject to a bomb threat and must be evacuated.
- **Hawaii Boating Excursion**: You are planning a trip among the Hawaiian Islands with your friend.
- **Bicycle Trip**: You are meeting up with your friend for a bicycle day-trip in the Minneapolis area.
- **Hiking Trail Addition**: You are a Park Ranger at the Grand Teton National Park in Wyoming. You and a colleague have been asked by your Head Ranger to plan an addition to the hiking trail near Mount Sheridan.
- **Mall Run**: You are at the mall with your friend and have to visit several stores quickly before the mall closes.

The full instructions are attached in Appendix I.

5.4 Questionnaire

The entire questionnaire includes two parts—the post-task questionnaire and the concluding questionnaire.

- The post-task questionnaire was used after each trial experimental condition and included two parts:

1) NASA Task Load Index (NASA TLX) for the six dependent variables: mental, physical and temporal demand, performance failure, effort, and frustration. NASA-TLX is a subjective, multidimensional assessment tool that rates the perceived workload, to access a task, a system, or a team's effectiveness or other aspects of performance. These statistical analyses have a prominent place in the history of Human Factors research.
2) A simple, yet specific, self-created questionnaire to access the collaborators' feelings of team cohesion, joint effort, social connection, team cooperation, and ease of communication in using the interface..

- The self-created concluding questionnaire was administered after running the experiments under five conditions. This concluding questionnaire is used to assess the differences between the five interfaces—to identify which interface conditions are considered to be the easiest, the most enjoyable, the most likely to be chosen for professional work, the favorite and least favorite, and the interface that made the participants feel most connected.

5.5 Procedure

The procedures are described as below. All sessions were videotaped.

1) Subject pairs were trained with a short practice session using the tool.
2) Subject pairs jointly solved five map-based planning problems. Each task was completed using different conditions. The task, condition, and presentation order were systematically varied. The process of completing the tasks was video recorded.
3) After each experiment under each condition, participants completed a post-task questionnaire.
4) After all the experiments under five conditions, participants completed a concluding questionnaire.
5) After the entire questionnaire had been completed, participants were invited for a short interview that was audio recorded. During the interviews, participants were asked how they felt about the experiment overall and how they felt about each of the five interface conditions. If there were any unexpected questionnaire results, the participants were asked to explain further.

5.6 Data Collected

The following data were collected to make analysis easier to assess whether or not the interface had changed the way in which participants communicated information to each other:

- **Task-completion time** of each trial to measure the performance;
- **Post-Task Questionnaire Data**, collected after one participant completed a task under a condition, including NASA's TXL and Self-created questions:
    1) **NASA's TXL Task Index Load Index post-task questionnaire** (which is a tool that finds out the cognitive workload involved in different scenarios [42]) to measure the workload;
    2) **Self-created post-task questionnaire** to measure the team cohesion;
- **Concluding questionnaire, along with the** data, collected after each participant completed all five tasks under the five different conditions to measure the preferences;
- **Open-ended interview**.

6. Results

In the results analysis, we collected data from task-completion times, post-task questionnaire results, concluding questionnaire results, and then commanded analyses separately.

First, for the task-completion time and the post-task questionnaire results, I commanded ANOVA analysis using ±95% confidence interval ("±95% confidence interval" means that the comparison pair, whose p-value is smaller than 0.05, indicates a significant difference) and pairwise comparisons. In this section, only the analyses for the all the significant results will be presented using a mean summary figure and a pairwise comparison table, followed by a short summary and discussion. The outputs of all the statistical analyses, whether significant or not, are included in Appendix III. Significant differences were found only for dependent variables:

1. Task-completion time (*extremely significant*)
2. Post-Task Questionnaire
   - NASA TLX
     1) Mental Demand
     6) Frustration *(marginally significant)*
   - Short Questions
     2) Connection Perspective: To what degree did you feel disconnected from your teammate?
     3) Interface Perspective: How well do you feel the computer interface allowed you to collaborate seamlessly with your partner?
     6) Solution: To what extent do you feel like you and your partner simply worked out your own individual solution to the problem and were only able to compare notes afterwards? (*significant in pairwise comparisons*)
     7) Connection: How much did you feel as if your partner was present with you, while working together on a solution?

Secondly, for the concluding questionnaire results, I made histograms to present the summary of those results and made a short summary and discussion for each result.

6.1 Task-Completion Time Analysis

For the time to complete a task, there was a significant main effect of interface, $p = 0.0098$, $F = 3.54$. Figure 18 on the next page shows a graphical comparison of the mean differences.

Furthermore, from the pairwise comparisons shown in Table 2 on the next page, we found the following three significantly different pairs whose p-values were lower than 0.05:

- the Gesture-Voice condition vs. the Voice-only condition ($p = 0.016$),
- the Mark-Gesture-Voice condition vs. the Voice-only condition ($p = 0.018$),
- the Face-to-Face condition vs. the Voice-only condition ($p = 0.022$).

Figure 18 confirms that, while the participants were working on the tasks, they required significantly less time in

- the Gesture-Voice condition (mean = 8.41) than in the Voice-only condition (mean = 13.82);
- the Mark-Gesture-Voice condition (mean = 8.8) than in the Voice-only condition (mean = 13.82);
- the Face-to-Face condition (mean = 8.72) than in the Voice-only condition (mean = 13.82).

(*To be continued on the next page)*

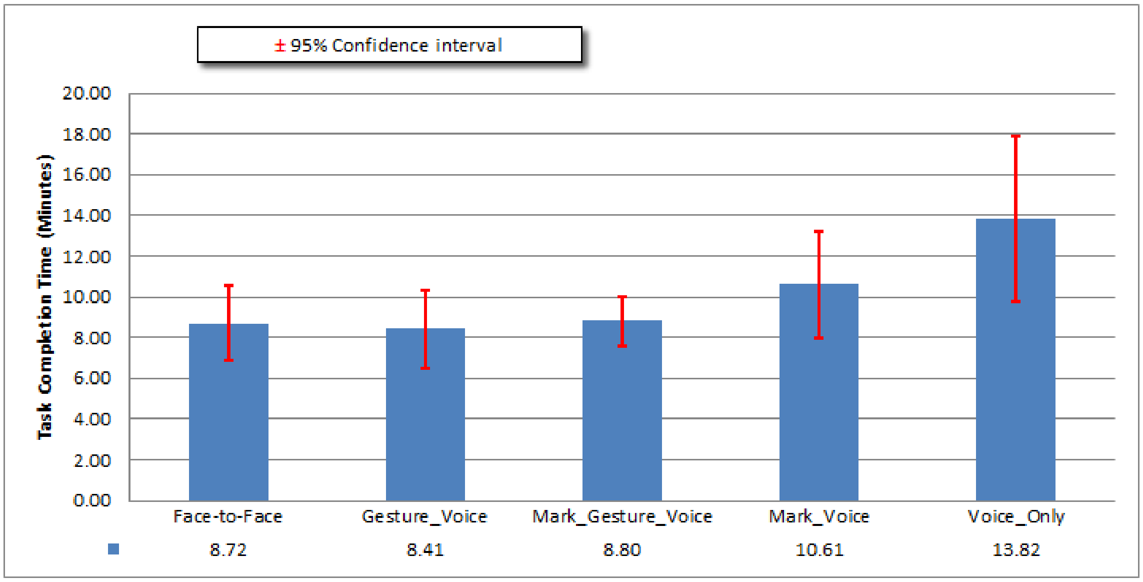


Figure 18: The Mean Task-completion time (measured by minutes). Error bars show the 95% confidence interval for each mean.

Task-completion time					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture Voice	0.956	0.956	0.059	0.810
Face-to-Face	Mark_Gesture_Voice	0.068	0.068	0.006	0.937
Face-to-Face	Mark_Voice	35.942	35.942	1.558	0.220
Face-to-Face	**Voice_Only**	**260.355**	**260.355**	**5.734**	**** 0.022***
Gesture_Voice	Mark_Gesture_Voice	1.534	1.534	0.129	0.721
Gesture_Voice	Mark_Voice	48.620	48.620	2.017	0.164
Gesture_Voice	**Voice_Only**	**292.861**	**292.861**	**6.305**	**** 0.016***
Mark_Gesture_Voice	Mark_Voice	32.882	32.882	1.767	0.192
Mark_Gesture_Voice	**Voice_Only**	**252.004**	**252.004**	**6.154**	**** 0.018***
Mark_Voice	Voice_Only	102.827	102.827	1.933	0.173

Table 2: Pairwise Comparison of Task-completion time

6.2 Post-Task Questionnaire Analysis

6.2.1 Part 1- NASA TLX Result

NASA Task Load Index (TLX) includes

1) Mental Demand *(significant)*,
2) Physical Demand,
3) Temporal Demand,
4) Performance,
5) Effort,
6) Frustration *(significant)*.

1) Mental Demand

For the NASA TLX measure of mental demand, there was a significant main effect of interface, $p = 0.024$, $F = 3.000$. Figure 19 on the next page shows a graphical comparison of the mean differences.

Also, from the pairwise comparisons shown in Table 3 on the next page, we found the following two significantly different pairs whose p-values were lower than 0.05:

- the Gesture-Voice condition vs. the Voice-only condition ($p = 0.011$),
- the Mark-Gesture-Voice condition vs. the Voice-only condition ($p = 0.05$).

This means that, while the participants were working on the tasks, they felt significantly less mental demand in

- the Gesture-Voice condition (mean = 2.7) than in the Voice-only condition (mean = 4.65);
- the Mark-Gesture-Voice condition (mean = 3.25) than in the Voice-only condition (mean = 4.65).

There is an unexpected result that the mental demand under the Face-to-Face condition was not significantly different from the Voice-only condition. During the interviews, participants stated that they did not feel ease under the Face-to-Face condition because they felt distracted when collaborating face-to-face with their partners. We might conclude that, although the Face-to-Face condition decreases communication barriers present in remote collaborations, the collaborators might feel extra mental demands, such as social anxiety, under the Face-to-Face condition.

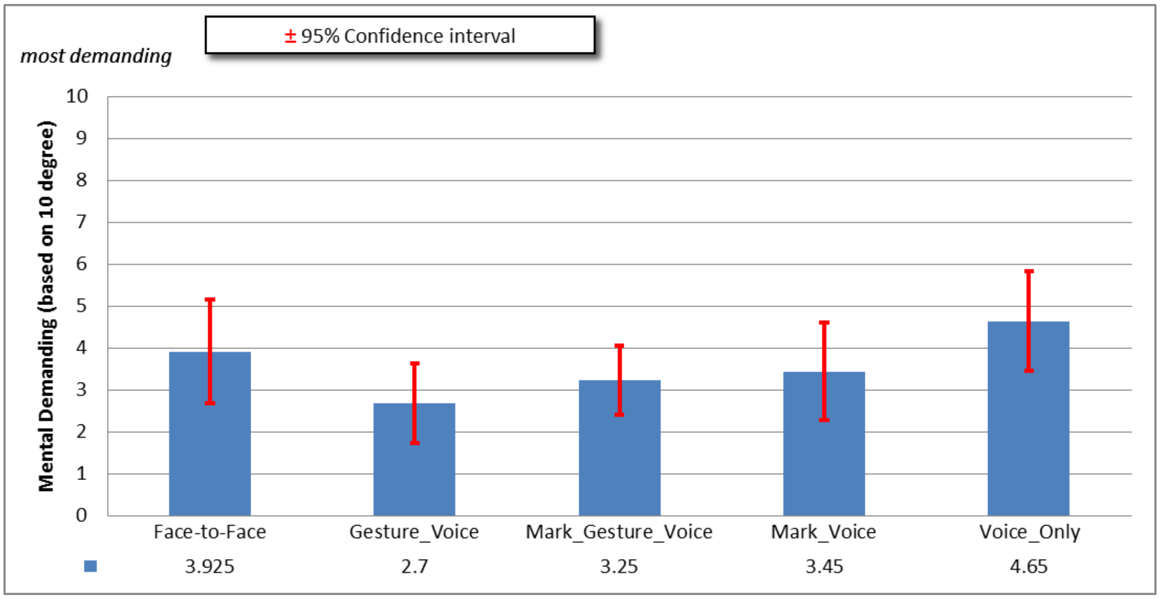


Figure 19: the Mean Mental Demand (based on a 10-point scale). Error bars show the 95% confidence interval for each mean.

Mental Demand					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	15.006	15.006	2.698	0.109
Face-to-Face	Mark_Gesture_Voice	4.556	4.556	0.900	0.349
Face-to-Face	Mark_Voice	2.256	2.256	0.342	0.562
Face-to-Face	Voice_Only	5.256	5.256	0.783	0.382
Gesture_Voice	Mark_Gesture_Voice	3.025	3.025	0.836	0.366
Gesture_Voice	Mark_Voice	5.625	5.625	1.093	0.303
Gesture_Voice	**Voice_Only**	**38.025**	**38.025**	**7.216**	*** 0.011**
Mark_Gesture_Voice	Mark_Voice	0.400	0.400	0.086	0.771
Mark_Gesture_Voice	**Voice_Only**	**19.600**	**19.600**	**4.108**	*** 0.050**
Mark_Voice	Voice_Only	14.400	14.400	2.285	0.139

Table 3: Pairwise Comparison of Mental Demand

2) Physical Demand

There was no significant difference between groups and pairs. The statistical tables and figures are attached in Appendix III.

3) Temporal Demand

There was no significant difference between groups and pairs. The statistical tables and figures are attached in Appendix III.

4) Performance

There was no significant difference between groups and pairs. The statistical tables and figures are attached in Appendix III.

5) Effort

There was no significant difference between groups and pairs. The summary tables and figures are attached in Appendix III.

6) Frustration

For the NASA-TLX measure of frustration level, there was a marginally significant effect of interfaces, $p = 0.05$, $F = 1.72$. Figure 20 on the next page shows a graphical comparison of the mean differences. Also, from the pairwise comparisons shown in Table 4 on the next page, we found one significantly different pair whose p-value was lower than 0.05: the Gesture-Voice condition vs. the Voice-only condition ($p = 0.018$). This means that participants perceived significantly more frustration under the Voice-only condition (mean = 2.725) than the Gesture-Voice condition (mean = 1.275).

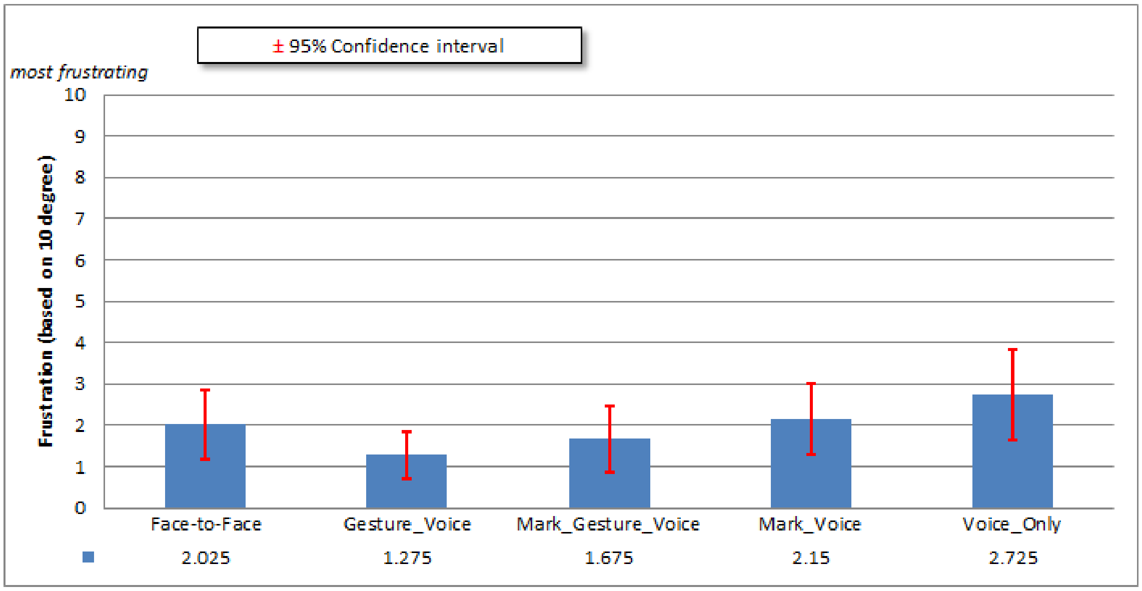


Figure 20: The Mean Frustration Level (based on a 10-point scale). Error bars show the 95% confidence interval for each mean.

Frustration Level					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	5.625	5.625	2.430	0.127
Face-to-Face	Mark_Gesture_Voice	1.225	1.225	0.400	0.531
Face-to-Face	Mark_Voice	0.156	0.156	0.047	0.829
Face-to-Face	Voice_Only	4.900	4.900	1.136	0.293
Gesture_Voice	Mark_Gesture_Voice	1.600	1.600	0.743	0.394
Gesture_Voice	Mark_Voice	7.656	7.656	3.170	0.083
Gesture_Voice	**Voice_Only**	**21.025**	**21.025**	**6.171**	**** 0.018***
Mark_Gesture_Voice	Mark_Voice	2.256	2.256	0.713	0.404
Mark_Gesture_Voice	Voice_Only	11.025	11.025	2.654	0.112
Mark_Voice	Voice_Only	3.306	3.306	0.749	0.392

Table 4: The Pairwise Comparison of Frustration Level

7) Aggregated NASA TLX

For the aggregated NASA TLX, there was no significant difference between groups and pairs. The summary tables and figures are attached in Appendix III.

6.2.2 Part 2- Short Questions

The second half of the post-task questionnaire is aimed at putting the analysis of solution, connection, interface, and cooperation as aspects under the different interface conditions. There are seven multiple-choice questions in all.

1) **Solution**: To what extent did your solution truly feel like a joint effort?
2) **Connection**: To what degree did you feel disconnected from your teammate? *(significant)*
3) **Interface**: How well do you feel the computer interface allowed you to collaborate seamlessly with your partner? *(significant)*
4) **Cooperation**: To what extent were you and your partner a team rather than two separate individuals?
5) **Interface**: To what degree did the computer interface limit your ability to collaborate with your partner?
6) **Solution**: To what extent do you feel like you and your partner simply worked on your own individual solutions to the problem, and were only able to compare notes afterwards? *(significant in pairwise comparisons)*
7) **Connection**: How much did you feel as if your partner was present with you, while working together on a solution? *(significant)*

1) Solution: To what extent did your solution truly feel like a joint effort?

There was no significant difference between groups and pairs. The statistical tables and figures are attached in Appendix III.

2) Connection: To what degree did you feel disconnected from your teammate?

When participants were asked, "To what degree did you feel disconnected from your teammate?" there was reported a significant main effect of the interface, p = 0.039, F = 2.63. The following figure shows a graphical comparison of the mean differences.

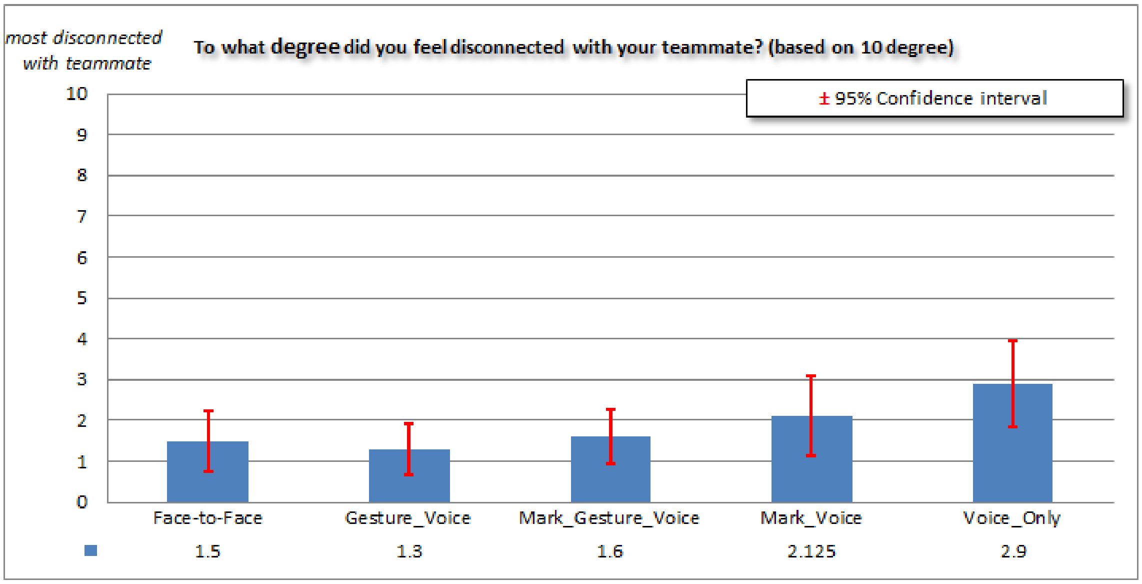


Figure 21: the Mean Disconnection Level (based on a 10-point scale). Error bars show the 95% confidence interval for each mean.

Connection: To what degree did you feel disconnected with your teammate?					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	0.400	0.400	0.184	0.671
Face-to-Face	Mark_Gesture_Voice	0.100	0.100	0.044	0.835
Face-to-Face	Mark_Voice	3.906	3.906	1.138	0.293
Face-to-Face	**Voice_Only**	**19.600**	**19.600**	**5.126**	**** 0.029***
Gesture_Voice	Mark_Gesture_Voice	0.900	0.900	0.472	0.496
Gesture_Voice	Mark_Voice	6.806	6.806	2.227	0.144
Gesture_Voice	**Voice_Only**	**25.600**	**25.600**	**7.426**	**** 0.010***
Mark_Gesture_Voice	Mark_Voice	2.756	2.756	0.871	0.357
Mark_Gesture_Voice	**Voice_Only**	**16.900**	**16.900**	**4.754**	**** 0.036***
Mark_Voice	Voice_Only	6.006	6.006	1.277	0.266

Table 5: The Pairwise Comparison of Disconnection Level

From the pairwise comparisons shown in the above table, we found the following significantly different pairs whose p-values were lower than 0.05:

- the Gesture-Voice condition vs. the Voice-only condition ($p = 0.010$);
- the Face-to-Face condition vs. the Voice-only condition ($p = 0.029$);
- the Mark-Gesture-Voice condition vs. the Voice-only condition ($p = 0.036$).

This means that the participants perceived significantly less disconnection from their teammates in

- the Gesture-Voice condition (mean = 1.3) than in the Voice-only condition (mean = 2.9);
- the Face-to-Face condition (mean = 1.5) than in the Voice-only condition (mean = 2.9);
- the Mark-Gesture-Voice condition (mean = 1.6) than in the Voice-only condition (mean = 2.9).

3) Interface: How well do you feel the computer interface allowed you to collaborate seamlessly with your partner?

When participants were asked, "How well do you feel the computer interface allowed you to collaborate seamlessly with your partner?" participants' answers are significantly different among interfaces, $p = 0.046$, $F = 2.553$. The figure on the next page shows a graphical comparison of the mean differences.

From the pairwise comparisons shown in Table 6 on the next page, we found one significantly different pair whose p-value was lower than 0.05: the Gesture-Voice condition vs. the Voice-only condition ($p = 0.002$). This means that participants had significantly more seamless collaborations with their partners under the Gesture-Voice condition (mean = 1.35) than the Voice-only condition (mean = 3.025). It is unexpected that the Face-to-Face condition was not shown to again be significantly different from the Voice-only condition. From the observations made during the experiments under the Face-to-Face condition, when the participants wanted to mark on the same area at the same time, since they only had one shared virtual workspace, they had to take turns to mark on the same area of the virtual map. This made the collaborators get hold sometime and feel less seamless. In contrast, under the remote conditions, when the participants wanted to mark on the same area at the same time, since they have separated interfaces, they could mark on the same area of the virtual map with no worries that their hands may overlap. This may be the reason that the participants did not feel the interaction was significantly more seamless when collaborating Face-to-Face versus Voice-only.

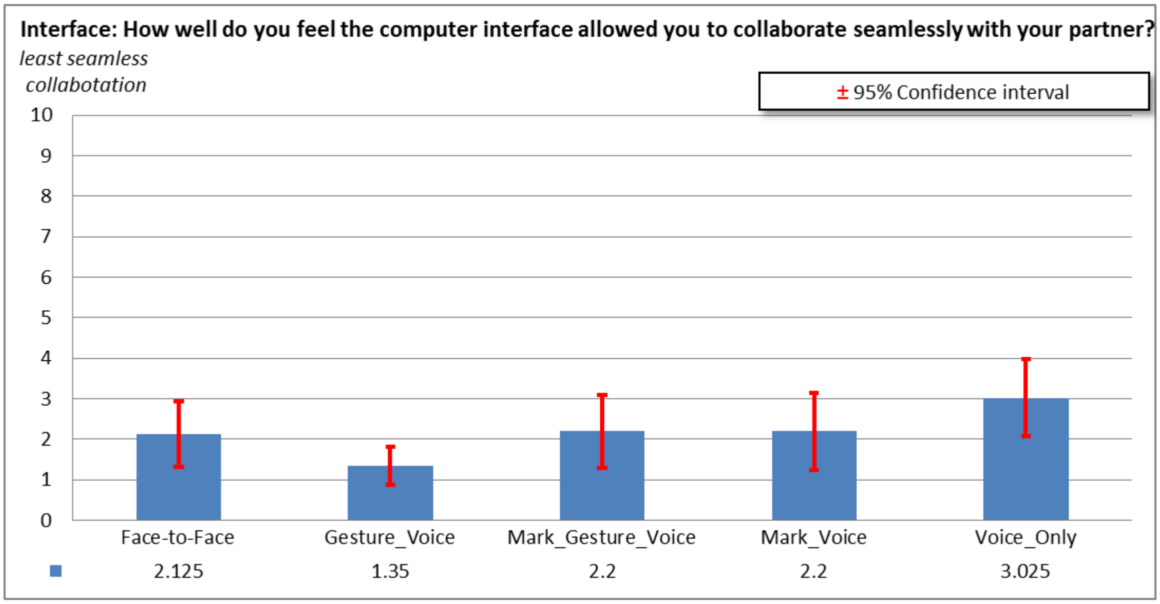


Figure 22: The Mean for Seamlessness-Levels of the Collaborative Interface (based on a 10-point scale). Error bars show the 95% confidence interval for each mean.

Interface: How well do you feel the computer interface allowed you to collaborate seamlessly with your partner?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	6.006	6.006	2.945	0.094
Face-to-Face	Mark_Gesture_Voice	0.056	0.056	0.017	0.898
Face-to-Face	Mark_Voice	0.056	0.056	0.016	0.901
Face-to-Face	Voice_Only	8.100	8.100	2.260	0.141
Gesture_Voice	Mark_Gesture_Voice	7.225	7.225	3.076	0.088
Gesture_Voice	Mark_Voice	7.225	7.225	2.809	0.102
Gesture_Voice	**Voice_Only**	**28.056**	**28.056**	**10.903**	**** 0.002***
Mark_Gesture_Voice	Mark_Voice	0.0001	0.0001	0.001	0.999
Mark_Gesture_Voice	Voice_Only	6.806	6.806	1.748	0.194
Mark_Voice	Voice_Only	6.806	6.806	1.653	0.206

Table 6: The Pairwise Comparison of Seamlessness-Levels of the Collaborative Interface

4) Cooperation: To what extent were you and your partner a team rather than two individuals?

There was no significant difference between groups and pairs. The statistical tables and figures are attached in Appendix III.

5) Interface: To what degree did the computer interface limit your ability to collaborate with your partner?

There was no significant difference between groups and pairs. The statistical tables and figures are attached in Appendix III.

6) Solution: To what extent do you feel like you and your partner simply worked your own individual solution to the problem, and were only able to compare notes afterwards?

When the participants were asked, "To what extent do you feel like you and your partner simply worked your own individual solution to the problem, and were only able to compare notes afterwards," although participants' answers are not significantly different among interfaces ($p = 0.0623$, $F = 2.231$), significant pairs were found in pairwise comparisons as shown in Table 7. The figure on the next page shows a graphical comparison of the mean differences.

From the pairwise comparisons shown in Table 6 on the next page, we found the following two significantly different pairs whose p-values were lower than 0.05:

- the Face-to-Face condition vs. the Voice-only condition ($p = 0.049$);
- the Face-to-Face condition vs. the Mark-Voice condition ($p = 0.024$).

This means the participant felt much less like an individual in

- the Face-to-Face condition (mean = 1.55) than in the Mark-Voice condition (mean = 2.7);
- the Face-to-Face condition (mean = 1.55) than in the Voice-only condition (mean = 2.85).

Those results indicate that when participants could not make gestures to each other, they felt significantly more like two individuals coordinating their separated solutions rather than two, creating one joint solution.

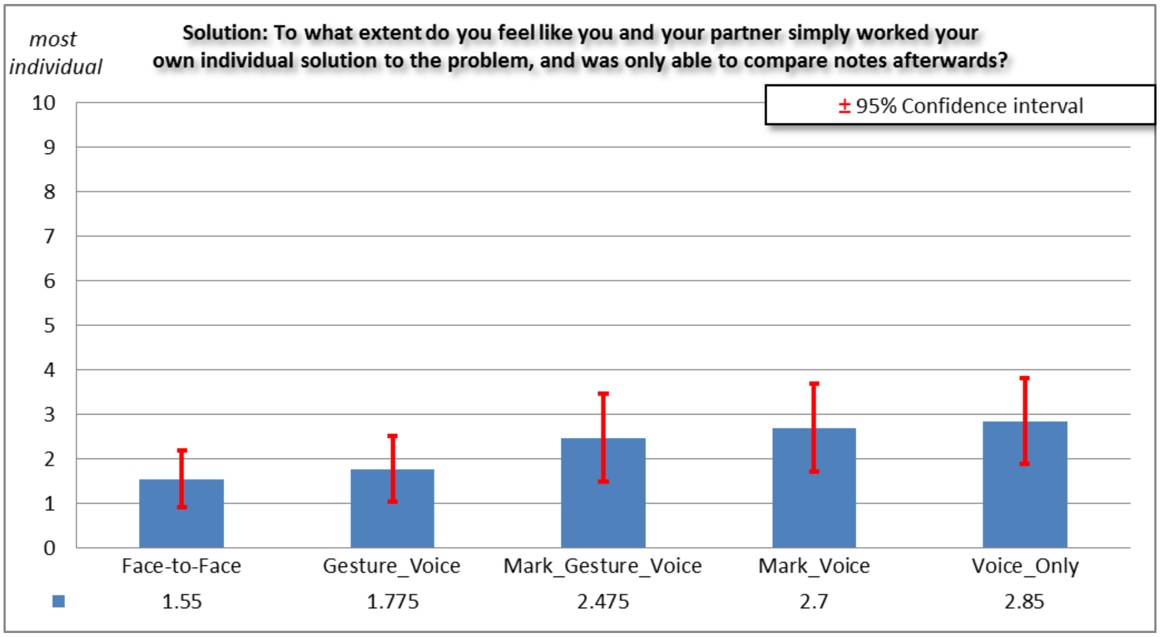


Figure 23: **The Mean Individual-Solution Level (based on a 10-point scale). Error bars show the 95% confidence interval for each mean.**

Solution: To what extent do you feel like you and your partner simply worked your own individual solution to the problem, and was only able to compare notes afterwards?					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	0.506	0.506	0.236	0.630
Face-to-Face	Mark_Gesture_Voice	8.556	8.556	2.739	0.106
Face-to-Face	**Mark_Voice**	**13.225**	**13.225**	**4.148**	***0.049**
Face-to-Face	**Voice_Only**	**16.9**	**16.9**	**5.560**	***0.024**
Gesture_Voice	Mark_Gesture_Voice	4.900	4.900	1.438	0.238
Gesture_Voice	Mark_Voice	8.556	8.556	2.464	0.125
Gesture_Voice	Voice_Only	11.556	11.556	3.477	0.070
Mark_Gesture_Voice	Mark_Voice	0.506	0.506	0.114	0.738
Mark_Gesture_Voice	Voice_Only	1.406	1.406	0.327	0.571
Mark_Voice	Voice_Only	0.225	0.225	0.052	0.822

Table 7: The Pairwise Comparison of Individual-Solution Level

7) Connection: How much did you feel as if your partner was present with you, while working together on a solution?

When the participants were asked, “How much did you feel as if your partner was present with you, while working together on a solution?” there was a significant main effect of interface, $p = 0.038$, $F = 4.225$. The figure on the next page shows a graphical comparison of the mean differences.

Also, from the pairwise comparisons shown in Table 8 on the next page, we found the following two significantly different pairs whose p-values were lower than 0.05:

- the Face-to-Face condition vs. the Voice-only condition ($p = 0.010$) ;
- the Face-to-Face condition vs. the Mark-Voice condition ($p = 0.013$).

This means the participant felt much more present in

- the Face-to-Face condition (mean = 0.95) than in the Mark-Voice condition (mean = 1.875);
- the Face-to-Face condition (mean = 0.95) than in the Voice-only condition (mean = 1.9).

These results indicate that when gestures were absent, participants have significantly less sense of their partners’ presence than under the Face-to-Face condition; shared maps and shared markings do not add to a sense of connection between virtual participants. It is only when gestures were added to the interface that the virtual interface became statistically indistinguishable from the Face-to-Face condition.

(*To be continued on the next page)*

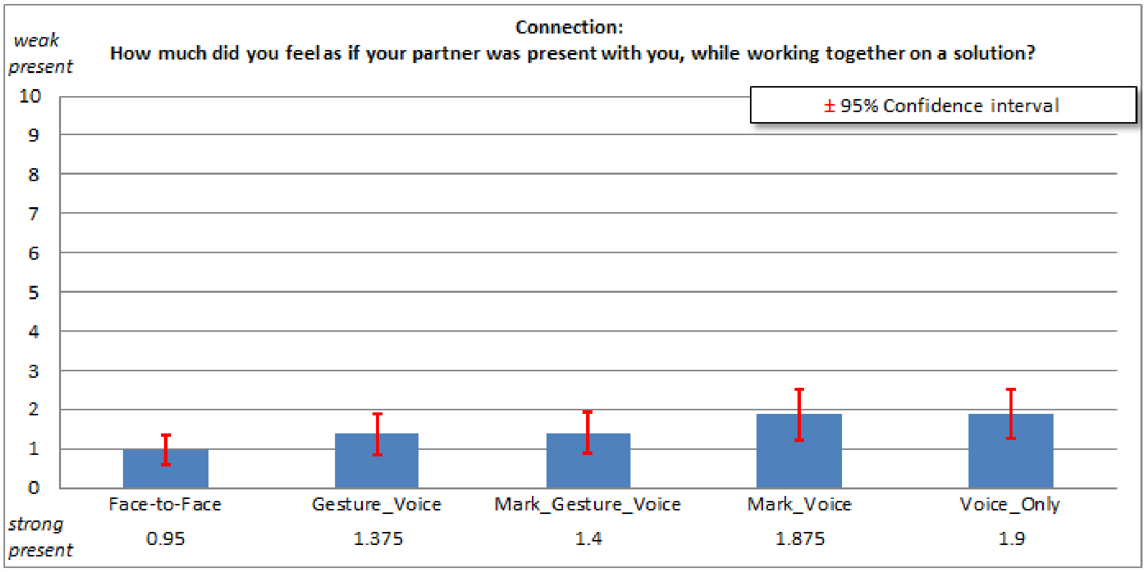


Figure 24: **The Mean Connection Level (based on a 10-point scale). Error bars show the 95% confidence interval for each mean.**

Connection: How much did you feel as if your partner was present with you, while working together on a solution?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	7.225	7.225	1.940	0.172
Face-to-Face	Mark_Gesture_Voice	8.1	8.1	2.247	0.142
Face-to-Face	**Mark_Voice**	**34.225**	**34.225**	**6.861**	****0.013***
Face-to-Face	**Voice_Only**	**36.1**	**36.1**	**7.415**	****0.010***
Gesture_Voice	Mark_Gesture_Voice	0.025	0.025	0.005	0.943
Gesture_Voice	Mark_Voice	10	10	1.614	0.212
Gesture_Voice	Voice_Only	11.025	11.025	1.814	0.186
Mark_Gesture_Voice	Mark_Voice	9.025	9.025	1.485	0.231
Mark_Gesture_Voice	Voice_Only	10	10	1.678	0.203
Mark_Voice	Voice_Only	0.025	0.025	0.003	0.954

Table 8: The Pairwise Comparison of Connection Level

6.3 Concluding Questionnaire Analysis

The concluding questionnaire is aimed to analyze the users' experience regarding the differences between the ease of use, enjoyableness, worthiness for working in a professional capacity, social connections, and the overall favorableness of the interface under different conditions. There are six multiple-choice questions in all:

1) I felt that all the computer interfaces were roughly the same in terms of usability, enjoyableness, and productivity. (Agree / Disagree)
2) Which interface was easiest to use? (Face-to-Face, Voice-only, Mark-Voice, Gesture-Voice, Mark-Gesture-Voice)
3) Which was the most fun? (Face-to-Face, Voice-only, Mark-Voice, Gesture-Voice, Mark-Gesture-Voice)
4) Which would you choose as a working professional collaborating with people in other locations? (Voice-only, Mark-Voice, Gesture-Voice, Mark-Gesture-Voice)
5) For which interface did you feel most connected to your partner? (Voice-only, Mark-Voice, Gesture-Voice, Mark-Gesture-Voice)
6) Please List the different computer interfaces from 1st favorite to least favorite: Face-to-Face, Voice-only, Mark-Voice, Gesture-Voice, and Mark-Gesture-Voice.

The results of the first question did not show that the participants experienced significant differences under five different interface conditions: half of the participants agreed that they felt that all the computer interfaces were roughly the same in terms of usability, enjoyableness, and productivity; half of the participants disagreed (shown in the following figure).

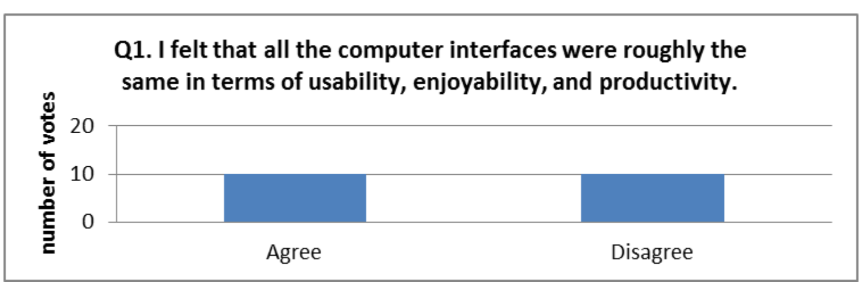


Figure 25: Feeling Difference among Five Different Interfaces (based on twenty participants)

The results of the second short question showed that, while the Face-to-Face condition was easier to use than all the remote collaborative conditions, among all the remote collaborative conditions, the Mark-Gesture-Voice condition was the easiest to use. See the Figure below.

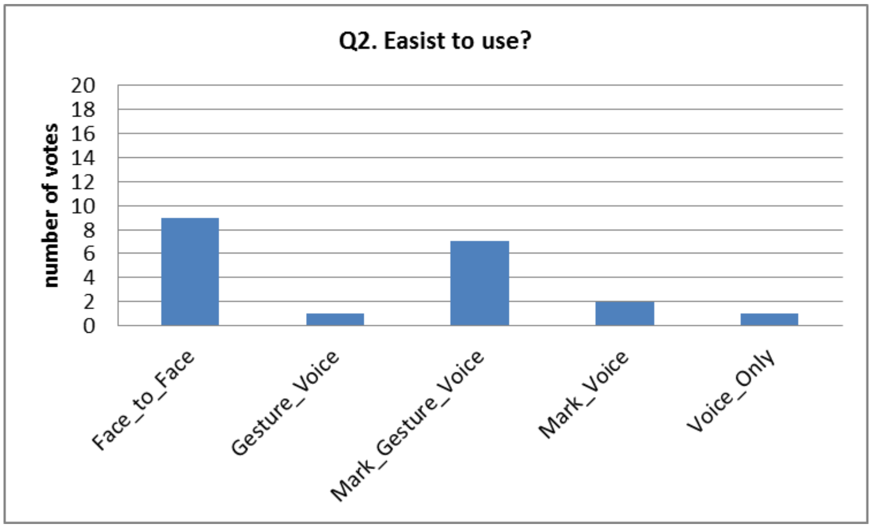


Figure 26: Easiest to Use Summary (based on twenty participants)

On the other hand, although the Face-to-Face condition was voted to be the easiest to use condition among all the conditions, it was not voted to be to the most enjoyable condition. The Mark-Gesture-Voice condition was voted as the most enjoyable condition. See the Figure below.

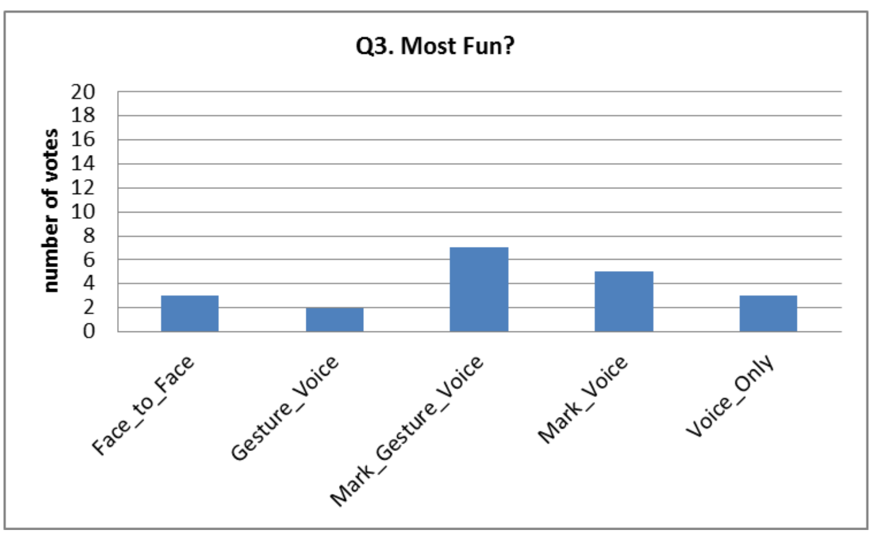


Figure 27: Most Enjoyable to Use Summary (based on twenty participants)

From the professional collaboration perspective, when the participants were asked which one they would use as a working professional collaborating with people in other locations, the Mark-Gesture-Voice condition was also chosen much more frequently than the others. Also, the participants expressed that they felt more connected under this condition than the others (for the question of the most connected interface, we also did not include the Face-to-Face condition as one of the answer choices. It is because that, in this question, we are more interested in finding out which remote interface help the collaborators feel most connected in remote collaborations). See the following two figures below.

Figure 28: Professional Work Preference Summary (based on twenty participants)

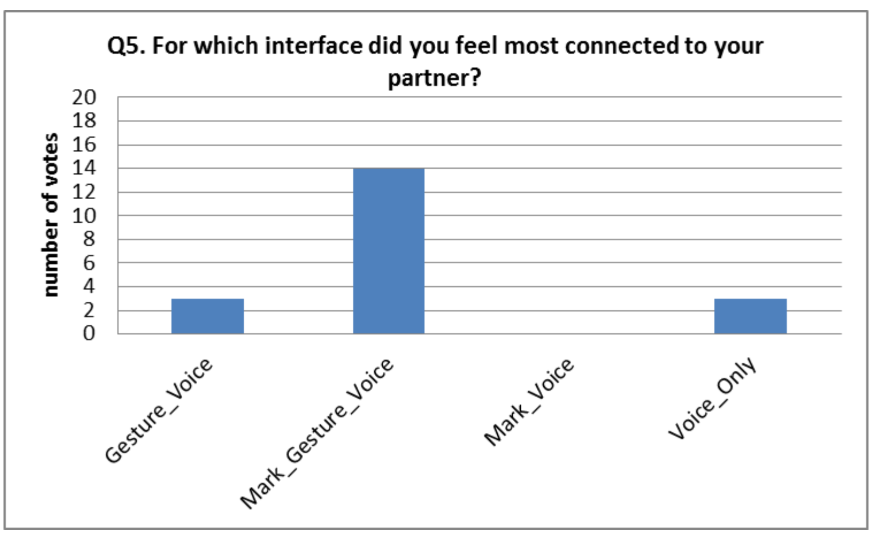


Figure 29: Connected Summary (based on twenty participants)

Regarding the least and most favorite, the Voice-only condition ranked as the least favorite (voted 10 times based on 20 participants); the Face-to-Face condition ranked as the most favorite (voted 8 times based on 20 participants) and the Mark-Gesture-Voice ranked as the second favorite (voted 5 times based on 20 participants). See the figures below.

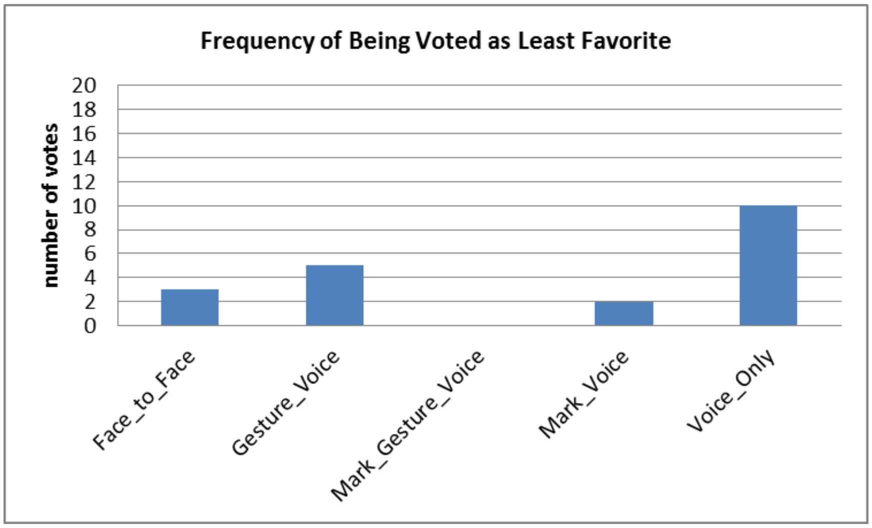


Figure 30: Frequency of "Least Favorite" Summary (based on twenty participants)

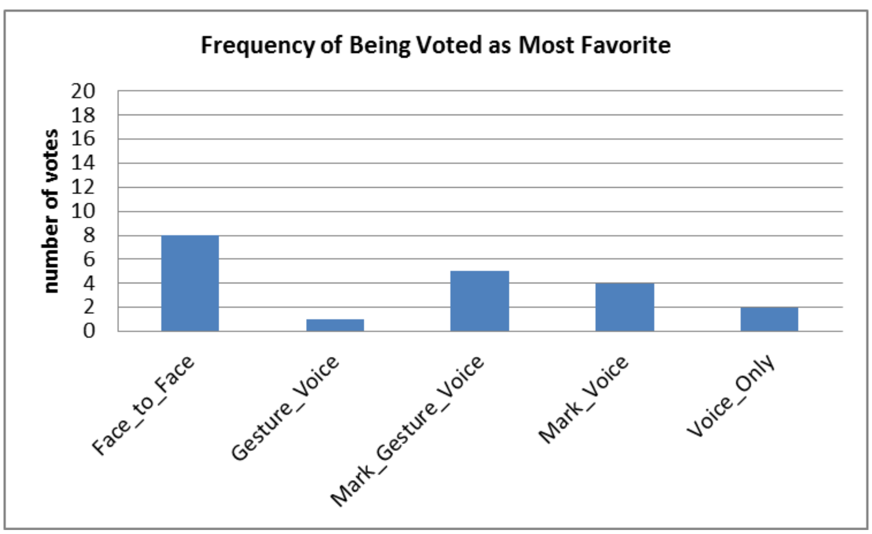


Figure 31: Frequency of "Most Favorite" Summary (based on twenty participants)

To summarize all the results:

Firstly, from performance perspective,

- **Task-Completion Time (extremely significant)**: participants spent significantly less task-completion time under the Face-to-Face condition, the Gesture-Voice condition and the Mark-Gesture-Voice condition than under the Voice-only condition;
- **Post-Task Questionnaire**

  <u>Part 1- NASA TLX</u>

  - **Mental Demand**: participants felt significantly less mental demands under the Gesture-Voice condition and the Mark-Gesture-Voice condition than under the Voice-only condition;
  - **Frustration (marginally significant):** participants felt significantly less frustrated under the Gesture-Voice condition than under the Voice-Only condition;

  <u>Part 2- Short Questions</u>

  - **Connection**: participants felt significantly less disconnected under the Gesture-Voice condition, the Face-to-Face condition, and the Mark-Gesture-Voice condition than under the Voice-only condition;
  - **Interface**: participants were able to collaborate significantly more seamlessly under the Gesture-Voice condition than under the Voice-Only condition;
  - **Solution**: participants felt significantly more like they were working separately on individual solutions under the Mark-Voice condition and the Voice-Only condition than under the Face-to-Face condition;
  - **Connection**: participants felt that their remote partner significantly was less present under the Mark-Voice condition and the Voice-Only condition relative to the Face-to-Face condition.

Secondly, from preference perspective,

- **Concluding Questionnaire**: <u>among all the remote conditions,</u> participants stated that they felt *the Mark-Gesture-Voice condition* is the

1) Easiest to use,
2) Most enjoyable,
3) Most likely to be chosen for professional work,
4) Most connected to the partner,
5) Favorite;

However, they still like *the Face-to-Face condition* better than any of remote conditions and felt it easiest to use among all the conditions, as shown in the graph below.

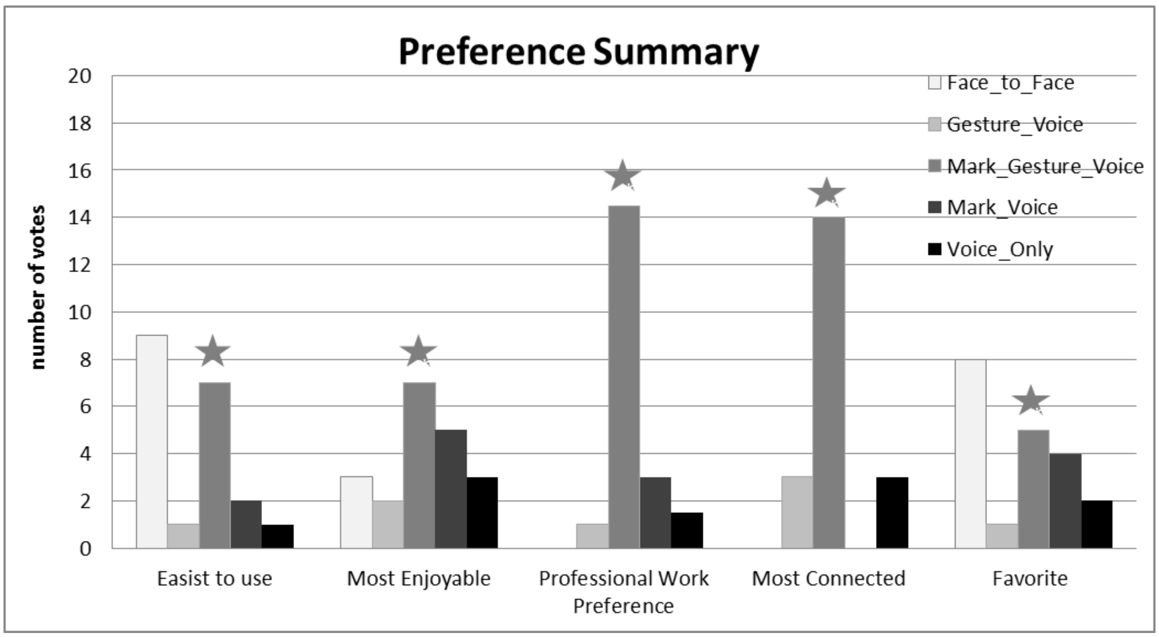


Figure 32: Summary of Concluding Questionnaire Based on Twenty Participants (Notes: the Face-to-Face condition is not voted in "Professional Work Preference" and "Most Connected" questions; the Star symbol indicates the condition which was voted most times among all the *remote* conditions.)

7. Discussion

In this research, we explored the extent to which hand images impact collaborative effectiveness in the task of map-based planning, with the goal of providing information on designing the tools that are needed to support virtual collaboration for map-based planning. This section will explain how our results go with the original hypothesis: By providing virtual collaborators a natural vehicle—hand images—performance and team cohesion will improve and workload will decrease. I will explain the results and also the surprises that were found. The following two points were concluded from the experimental results, and will be discussed in this section:

1) *When hand images were added to the virtual collaborative interface, the performance was significantly improved relative to the Voice-only condition; however, just adding a shared map and a shared marking did not result in significantly better performance than with the Voice-only interface.*
2) *Participants preferred the "richest" condition, Mark-Gesture-Voice, and felt it was most enjoyable.*

1) When hand images were added to the virtual collaborative interface, the performance was significantly improved relative to the Voice-only condition; however, just adding a shared map and a shared marking did not result in significantly better performance than with the Voice-only interface.

The results from the previous section indicated that, as compared with the Voice-only condition, adding gestures to the virtual interface (referred to the Gesture-Voice condition and the Mark-Gesture-Voice condition) could significantly improve interface performance in terms of task-completion time, workload (mental demand & frustration level), and team cohesion (connection level & collaboration seamlessness); however, only adding a shared map and a shared marking (referred to the Mark-Voice condition) did not result in significantly better performance.

From the observation of the experiments under the gesture-involved conditions, we saw that the addition of remote gestures to speech alleviated the likelihood of interruption when expressions were being formed or modified or when a speaker wished to retain the floor. Therefore, adding remote gestures to the interface reduced the number of failed attempts at turn-taking and helped collaborators spend significantly less time on costly sentence-repair phases. For example, when a collaborator wanted to point out a Hawaiian city (Napili-Honokowai) to his partner, he or she may not have been able to pronounce the island's name exactly or circle it quickly, but might have easily and quickly pointed out the place using gestures, thereby saving the collaborators the time of costly sentence repairs. However, under the Mark-Voice condition, we observed that collaborators nevertheless put lots of effort into changing and repairing their groundings. For example, they spent lots of time erasing their markings or re-marking. From the above results and observations, we can conclude that, in remote collaborations, the shared marking alone was not helpful enough to significantly improve the performance from the Voice-only condition; while, gesture-involved conditions were significantly helpful for improving the performance.

Kirk and Fraser [16] also demonstrated similar performance benefits from using hand-based gestures in collaborative physical tasks, showing that higher rates of remote-gesture use were correlated with faster task performance. This result is consistent with my study's result wherein gesture-involved conditions lead to quicker performance. In addition, their results also indicated the superiority of hand-based conditions (hand-image-only condition) in remote collaboration systems when compared with any kind of pen-based conditions (sketches-only condition and sketches & hand images condition). Although my result did not find any significant difference between Gesture-Voice condition and any marking-involved condition, we did find that the performance of the Mark-Voice condition was not significantly improved from the Voice-only condition.

These results suggest that the marking function alone is not sufficient enough to significantly increase the virtual team's cohesion and the design of virtual collaboration tools should consider first adding remote gestures before adding markings to improve the performance.

2) Participants preferred the "richest" condition, Mark-Gesture-Voice, and felt it was the most enjoyable among all the remote conditions. The results of section 6.3 (Figure 25—31) showed that participants felt the Mark-Gesture-Voice condition was 1) the easiest to use, 2) the most fun, and 3) made them feel the greatest connection with their partners, among all the remote conditions. The Mark-Gesture-Voice condition was also chosen as the favorite professional collaboration tool for distant collaborations. Participants chose the Mark-Gesture-Voice most often as their favorite, among all the remote conditions, and none of the participants chose the Mark-Gesture-Voice interface as their least favorite interface. From the interviews, participants explained that the major reason they chose the Mark-Gesture-Voice condition as their favorite was because adding the marking interface to the hand-image interface made the whole workspace seem more technological and professional. Similar preference results were found in Cornelius' project, also: most participants preferred the Virtual Sketches & Hand image condition in her experiments. Those results suggested that the design team should add both hand images and the marking functions to the virtual collaboration tools to make collaboration more enjoyable. However, an extra marking interface added to the hand-image interface may not really improve the performance on average, since the Mark-Gesture-Voice condition was not significantly better than the other conditions from our statistical analysis.

In conclusion, from the performance perspective, any condition that involves gestures (Gesture-Voice condition and Mark-Gesture-Voice condition) significantly improved collaborators' performance, relieved their workload, and increased their team cohesion over the Voice-only condition in distance collaborations (as conducted on the phone, through which the current map-based collaborations are often carried out). On the other hand, from the preference perspective, participants preferred the "richest" condition, Mark-Gesture-Voice, and felt it to be the most enjoyable among all the remote conditions.

Thus, we suggested that, in order to improve the performance, the designers of tools should add the hand-images' vehicle first, in developing the remote collaboration tools; and second, to improve the quality of the user-experience, should also add the marking functions to the hand-image interface.

8. Conclusions

From the above result analyses and the discussions, it is easy to come to the conclusion that the two virtual collaborative interfaces that involve natural gestures significantly improved distance team's performance, decreased workload, and increased team cohesion. Compared with the Voice-only condition, in which map-based collaborations are often carried out (e.g., by phone),

- The virtual conditions which include gestures (Gesture-Voice condition and Mark-Gesture-Voice condition) significantly
    1) Reduced task completion time,
    2) Decreased mental demand,
    3) Helped participants to feel more connected to their teammates;
- Additionally, when using the Gesture-Voice condition, participants
    1) Felt significantly less frustrated,
    2) Collaborated significantly more seamlessly.

Thus, Gesture-Voice helped the most often, followed by Mark-Gesture-Voice. These results indicate that the role of hand images is quite important. Hand images can greatly impact the feeling of connection and seamless collaboration with a distance teammate and can decrease the mental demand and frustration level in distance collaboration, greatly improving a teammate's engagement in the tasks, and making communication more effective, thus preserving task-completion time overall.

On the other hand, results from the analysis of the concluding preference questionnaire indicated that,

- Among all the remote conditions, the interface that combined Mark and Gesture was the
    1) Easiest to use,
    2) Most enjoyable,
    3) Most often chosen as a professional collaboration tool,
    4) One that helped users feel most connected to the partner,
    5) Favorite;

- However, participants still like the Face-to-Face condition better than any of remote conditions and felt it easiest to use among all the conditions.

These results suggest that in order improve performance in joint map-planning tasks, designers of remote collaboration tools should consider adding technologies that allow users to communicate to remote partners using natural embedded gestures (e.g. videos of hands and arms overlaid on a shared virtual work surface). Additionally, to improve the appeal and enjoyment factor for the users, software designers should also consider adding the capability for users to make markings on the shared work surface that can be viewed by all.

9. Future Work

Future investigations for system interface improvements Include:

- From the perspective of the flexibility of location, the current system is cumbersome and not easily moved around, so this system is only more helpful to office workers or home users who are in fixed locations. A mobile version might be useful.
- From the perspective of ease of use, we used an infrared pen as the cursor of interaction because of cost. However, as touchscreen technology continues to grow and become cheaper than ever, a future interface could include a touch screen; this would not only improve the usability of the interaction, but also decrease the frustration in drawing or marking.
- From the perspective of user cohesiveness, two extra screens showing the participants' faces to each other could be added, so as to add the aspect of evaluation for the importance of facial cues in such interactions. However, such monitors could also disturb the participants, and thereby decrease the efficiency of their distance collaboration. This caveat is based on the current project's result showing that Face-to-Face collaboration was usually less efficient than the Gesture-Voice collaboration, on average because participants were more focused on their own workflow when separated from each other.

In addition to changing the tools, in the future, we could design additional scenarios to see how the system performs in a wider variety and complexity of tasks. We could also modify the system in order to study the impact of gestures in other task domains, for example, education and medicine. Additionally, we could develop this server-client communication mode into a group communication mode.

References

[1] Russo, Enio Emanuel Ramos, Alberto Raposo, Terrence Fernando, Marcelo Gattass, and Börje Karlsson. "CONFIGURING A COLLABORATIVE VIRTUAL WORKSPACE FOR DISASTER MANAGEMENT OF OIL & GAS OFFSHORE STRUCTURES." In *Joint International Conference on Computing and Decision Making in Civil and Building EngineeringEds*. 2006.

[2] Veinott, Elizabeth S., Judith Olson, Gary M. Olson, and Xiaolan Fu. "Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other." In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 302-309. ACM, 1999.

[3] Kraut, Robert E., Susan R. Fussell, and Jane Siegel. "Visual information as a conversational resource in collaborative physical tasks." *Human-computer interaction* 18, no. 1 (2003): 13-49.

[4] Ochsman, Robert B., and Alphonse Chapanis. "The effects of 10 communication modes on the behavior of teams during co-operative problem-solving." *International Journal of Man-Machine Studies* 6, no. 5 (1974): 579-619.

[5] Tang, John C., and Scott L. Minneman. "VideoDraw: a video interface for collaborative drawing." *ACM Transactions on Information Systems (TOIS)* 9, no. 2 (1991): 170-184.

[6] Tang, John C., and Scott Minneman. "VideoWhiteboard: video shadows to support remote collaboration." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 315-322. ACM, 1991.

[7] Xia, Steven, David Sun, Chengzheng Sun, David Chen, and Haifeng Shen. "Leveraging single-user applications for multi-user collaboration: the coword approach." In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 162-171. ACM, 2004.

[8] Fussell, Susan R., Leslie D. Setlock, Jie Yang, Jiazhi Ou, Elizabeth Mauer, and Adam DI Kramer. "Gestures over video streams to support remote collaboration on physical tasks." *Human-Computer Interaction* 19, no. 3 (2004): 273-309.

[9] Bly, Sara A., and Scott L. Minneman. "Commune: A shared drawing surface."*ACM SIGOIS Bulletin* 11, no. 2-3 (1990): 184-192.

[10] Drew, Daniel, Caroline C. Hayes, Mai-Anh Nguyen, and Xuan Cheng. "Hand Images in Virtual Spatial Collaboration for Traffic Incident and Disaster Management." (2013).

[11] Tang, Anthony, Michael Boyle, and Saul Greenberg. "Display and presence disparity in Mixed Presence Groupware." In *Proceedings of the fifth conference on Australasian user interface-Volume 28*, pp. 73-82. Australian Computer Society, Inc., 2004.

[12] Ishii, Hiroshi, and Minoru Kobayashi. "Clearboard: A seamless medium for shared drawing and conversation with eye contact." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 525-532. ACM, 1992.

[13] Ou, Jiazhi, Susan R. Fussell, Xilin Chen, Leslie D. Setlock, and Jie Yang. "Gestural communication over video stream: supporting multimodal interaction for remote collaborative physical tasks." In *Proceedings of the 5th international conference on Multimodal interfaces*, pp. 242-249. ACM, 2003.

[14] Kuzuoka, Hideaki, Jun'ichi Kosaka, Keiichi Yamazaki, Yasuko Suga, Akiko Yamazaki, Paul Luff, and Christian Heath. "Mediating dual ecologies." In*Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 477-486. ACM, 2004.

[15] Sakata, Nobuchika, Takeshi Kurata, Takekazu Kato, Masakatsu Kourogi, and Hideaki Kuzuoka. "WACL: Supporting telecommunications using wearable active camera with laser pointer." In *2012 16th International Symposium on Wearable Computers*, pp. 53-53. IEEE Computer Society, 2003.

[16] Kirk, David, and Danae Stanton Fraser. "Comparing remote gesture technologies for supporting collaborative physical tasks." In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1191-1200. ACM, 2006.

[17] Clark, Herbert H., and Meredyth A. Krych. "Speaking while monitoring addressees for understanding." *Journal of Memory and Language* 50, no. 1 (2004): 62-81.

[18] Kraut, Robert E., Susan R. Fussell, and Jane Siegel. "Visual information as a conversational resource in collaborative physical tasks." *Human-computer interaction* 18, no. 1 (2003): 13-49.

[19] Gergle, Darren, Robert E. Kraut, and Susan R. Fussell. "Action as language in a shared visual space." In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pp. 487-496. ACM, 2004.

[20] Fussell, Susan R., Leslie D. Setlock, and Elizabeth M. Parker. "Where do helpers look?: gaze targets during collaborative physical tasks." In *CHI'03 Extended Abstracts on Human Factors in Computing Systems*, pp. 768-769. ACM, 2003.

[21] Clark, Herbert H., and Susan E. Brennan. "Grounding in communication."*Perspectives on socially shared cognition* 13, no. 1991 (1991): 127-149.

[22] Luff, Paul, Christian Heath, Hideaki Kuzuoka, Keiichi Yamazaki, and Jun Yamashita. "Handling documents and discriminating objects in hybrid spaces." In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 561-570. ACM, 2006.

[23] McNeill, David. *Hand and mind: What gestures reveal about thought*. University of Chicago Press, 1992.

[24] Kramer, Adam DI, Lui Min Oh, and Susan R. Fussell. "Using linguistic features to measure presence in computer-mediated communication." In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 913-916. ACM, 2006.

[25] Krauss, Robert M., and Susan R. Fussell. "Constructing shared communicative environments." (1991).

[26] Schremmer, Claudia, and Christian Müller-Tomfelde. "HxI: research down under in distributed intense collaboration between teams." In *CHI'08 Extended Abstracts on Human Factors in Computing Systems*, pp. 3645-3650. ACM, 2008.

[27] McNelley, Steve. "Immersive Group Telepresence and the Perception of Eye Contact." (2005): 7.

[28] Pauchet, Alexandre, François Coldefy, Liv Lefebvre, S. Louis Dit Picard, Arnaud Bouguet, Laurence Perron, Joël Guerin, Daniel Corvaisier, and Michel Collobert. "Mutual awareness in collocated and distant collaborative tasks using shared interfaces." In *Human-Computer Interaction–INTERACT 2007*, pp. 59-73. Springer Berlin Heidelberg, 2007.

[29] Tang, Anthony, and Saul Greenberg. "Supporting awareness in mixed presence groupware." In *From Awareness to HCI Education: The CHI'2005 Workshop Papers Suite*, p. 10. 2005.

[30] Tang, Anthony, Michael Boyle, and Saul Greenberg. "Display and presence disparity in Mixed Presence Groupware." In *Proceedings of the fifth conference on Australasian user interface-Volume 28*, pp. 73-82. Australian Computer Society, Inc., 2004.

[31] Robinson, Peter, and Philip Tuddenham. "Distributed tabletops: Supporting remote and mixed-presence tabletop collaboration." In *Horizontal Interactive Human-Computer Systems, 2007. TABLETOP'07. Second Annual IEEE International Workshop on*, pp. 19-26. IEEE, 2007.

[32] Nam, Tek-Jin, and Kyung Sakong. "Collaborative 3D workspace and interaction techniques for synchronous distributed product design reviews." *International Journal of Design* 3, no. 1 (2009): 43-55.

[33] Genest, Aaron, and Carl Gutwin. "Evaluating the effectiveness of height visualizations for improving gestural communication at distributed tabletops." In*Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pp. 519-528. ACM, 2012.

[34] Wickey, Aiden, and Leila Alem. "Analysis of hand gestures in remote collaboration: some design recommendations." In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, pp. 87-93. ACM, 2007.

[35] Lee, Min Kyung, and Leila Takayama. "Now, I have a body: Uses and social norms for mobile remote presence in the workplace." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 33-42. ACM, 2011.

[36] Ou, Jiazhi, Lui Min Oh, Jie Yang, and Susan R. Fussell. "Effects of task properties, partner actions, and message content on eye gaze patterns in a collaborative task." In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 231-240. ACM, 2005.

[37] Cornelius, Caroline Joyce. "Development and Evaluation of Gesture Rich Collaborative Drawing tools." PhD diss., UNIVERSITY OF MINNESOTA, 2011.

[38] McDonald, Chris, Gerhard Roth, and Steve Marsh. "Red-handed: collaborative gesture interaction with a projection table." In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pp. 773-778. IEEE, 2004.

[39] Liao, Wen-Hung, and Yu-Hao Chi. "Estimation of Skin Color Range Using Achromatic Features." In *Intelligent Systems Design and Applications, 2008. ISDA'08. Eighth International Conference on*, vol. 2, pp. 493-497. IEEE, 2008.

[40] Desa, Shahbe Mat, and Qussay A. Salih. "Image subtraction for real time moving object extraction." In *Computer Graphics, Imaging and Visualization, 2004. CGIV 2004. Proceedings. International Conference on*, pp. 41-45. IEEE, 2004.

[41] D'Souza, Melroy E., and Joel S. Greenstein. "Listening to users in a manufacturing organization: a context-based approach to the development of a computer-supported collaborative work system." *International journal of industrial ergonomics* 32, no. 4 (2003): 251-264.

[42] Emgu, C. V. "Emgu cv." *URL http://emgucv. sourceforge. net* (2011).

[43] O'Conaill, Brid, and Steve Whittaker. "Characterizing, predicting, and measuring video-mediated communication: A conversational approach." *Video-mediated communication* (1997): 107-132.

[44] Delgado, Miguel A., and Peter M. Robinson. "Optimal spectral bandwidth for long memory." (1996).

[45] Clark, Herbert H. *Using language*. Vol. 1996. Cambridge: Cambridge University Press, 1996.

[46] Nakano, Yukiko I., Gabe Reinstein, Tom Stocky, and Justine Cassell. "Towards a model of face-to-face grounding." In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*, pp. 553-561. Association for Computational Linguistics, 2003.

Appendix

I. Task Description

General information for use in all scenarios

In these scenarios, each of you will be given different information. You will need to discuss that information in order to complete the tasks. Use the collaborative interface as needed to share information or discuss ideas. Jointly decide (1) what to do and (2) why.

Each scenario is designed to take approximately 10-15 minutes, though you are free to complete them sooner if you and your partner feel you have finished the task satisfactorily.

Tips for using the collaborative interface effectively:

- Go slowly. The system can’t process too many changes at once.
- Try to avoid drawing at the same time as your partner.
- For the straight line drawing tool, click only the endpoints. Use the curved line tool as you would draw with a pen.
- The erase function erases everything!
- Unfortunately, there is no button to undo a single action.
- Hold the infrared pen at an angle so that your hand does not obstruct the LED at the tip of the pen. The LED must face upwards to be seen by the system.

Task#1 Bomb Threat/Evacuation: Participant I

Situation: The IDS Center in downtown Minneapolis is subject to a bomb threat and must be evacuated.

You have: Occupancy information on the IDS Center itself, and the surrounding buildings. You quickly triage and determine that these locations are highest priority for evacuation:

Location to Evacuate	Current Occupancy
IDS Center	3,000
33 South Sixth /Marriott Hotel City Center	6,000
Plaza VII	4,000
Wells Fargo Center	6,000

Your coworker has: quickly gathered information about *evacuation safe zones* and hopes that you have enough capacity to house the people until the emergency is averted.

Planning Goals: Confer with your co-worker and come up with a solution to evacuate all the civilians in the area.

Show or tell your plan to the experimenter to complete the task.

Task#1 Bomb Threat/Evacuation: Participant II

Situation: The IDS Center in downtown Minneapolis is subject to a bomb threat and must be evacuated.

Your coworker has: quickly determined which buildings need to be evacuated, and their current rough occupancy.

You have: gathered a list of possible evacuation safe zones and their rough capacities:

Evacuation Safe Zone	Evac. Direction	Capacity
Metrodome (S 6th St & Chicago Ave S)	East	9,000
Target Field (N 7th St & 3rd Ave N)	Northwest	5,000
Mpls Convention Center (12th St & 2nd Ave S)	South	6,000

Planning Goals: Confer with your co-worker and come up with a solution to evacuate all the civilians in the area.

Show or tell your plan to the experimenter to complete the task.

Task#2 Hawaii Boating Excursion: Participant I

Situation: You are planning a trip among the Hawaii Islands with your friend.

Starting position:**Kalaupapa National Historical Park**

Places to Visit: Ask your friend what their ideas are.

Stipulations on places to visit:

1. There is a very recent oil leakin the area; the whole **waterway is closed** between the islands of *Lana'I* and *Maui* (the city of Lahaina is on its west coast) to contain the spill.
2. Ask your friend where they would like to visit.

Planning Goals:

Agree on what path to take for your trip, taking the closed-off area into consideration.

Show or tell your plan to the experimenter to complete the task.

Task#2 Hawaii Boating Excursion: Participant II

Situation: You are planning a trip among the Hawaii Islands with your friend.

Starting position:**Kalaupapa National Historical Park**

The sites you want to visit:

1. **Maalaea Bay**
2. **Ahihi Kinau Natural Area Reservation**
3. However, you heard about an oil leak, and that the coast guard had closed the affected area off to boating traffic. Luckily, your friend mentioned beforehand that he/she knows where it's closed.

Planning Goals:

Agree on what path to take for your trip, taking the closed-off area into consideration.

Show or tell your plan to the experimenter to complete the task.

Task#3 Bicycle Trip: Participant I

Situation: You are meeting up with your friend to for a bicycle day trip in the Minneapolis area.

Starting position: Downtown Minneapolis, **Park Ave S. and S 6th St.**

The sites you want to visit:The waterfall in **Minnehaha Park**; ask your friend what sites he/she would like to visit.

Planning Goals:

1. Agree upon where to meet
2. What path to take for your trip

Show or tell these to the experimenter to complete the task.

Task#3 Bicycle Trip: Participant II

Situation: You are meeting up with your friend to for a bicycle day trip in the Minneapolis area.

Starting position: The **Mechanical Engineering** building, on the U of M East Bank (pt.A on map)

The sites you want to visit:The path around**Lake Calhoun**; ask your friend what sites he/she would like to visit.

Planning Goals:

1. Agree upon where to meet
2. What path to take for your trip

Show or tell these to the experimenter to complete the task.

Task#4 Hiking Trail Addition: Participant I

Situation: You are a Park Ranger at the Grand Teton National Park in Wyoming. You and a colleague have been asked by your Head Ranger to plan an addition to the hiking trail near Mt Sheridan.

Starting Position:The new trail will branch from the existing network at a new trail head station marked with a red X on the map.

New Trail Points of Interest:

1. **Factory Hill**
2. The peak of **Mt. Sheridan**
3. Mt. Sheridan's southwestern valley, as that area is particularly beautiful in the fall
4. Ask your partner about his/her ideas as well.

For clarity, the existing trail is the dotted white line. The thin grey lines are the constant elevation topographical lines, and are not trails.

Planning Goals:

1. Agree on the path for the new section of the hiking trail.
2. Where the trail should connect back to the existing trail.

Show or tell your plan to the experimenter to complete the task.

Task#4 Hiking Trail Addition: Participant II

Situation: You are a Park Ranger at the Grand Teton National Park in Wyoming. You and a colleague have been asked by your Head Ranger to plan an addition to the hiking trail near Mt Sheridan.

Starting Position:The new trail will branch from the existing network at a new trail head station marked with a red X on the map.

The way to measure is to connect the new trail back to either a new or existing part of the trail, and measure the whole circuit starting and ending at the trail head. *Don't worry too much about the exact length numbers*, just treat them as a rough benchmark to aim for.

New Trail Points of Interest:

1. Aster Lake
2. Ask your partner about their ideas

For clarity, the existing trail is the dotted white line. The thin grey lines are the constant elevation topographical lines, and are not trails.

Planning Goals:

1. Agree on the path for the new section of the hiking trail.
2. Where the trail should connect back to the existing trail.

Show or tell your plan to the experimenter to complete the task.

Task#5 Mall Run: Participant I

Situation: You are at the mall with your friend, and have to visit several stores quickly before the mall closes.

Starting position: Sea Life (Underwater World, pt. marked on map)

The sites you want to visit:

1. Macy's
2. DSW Designer Shoe Warehouse
3. Hot Topic
4. Ask your friend where they need to go as well.

Planning Goals:

1. Agree upon the quickest path to reach all the stores
2. Estimate the (very) rough time it will take to visit all the places on your list, if you spend 5 minutes at each store.
3. Show or tell these to the experimenter to complete the task.

Task#5 Mall Run: Participant II

Situation: You are at the mall with your friend, and have to visit several stores quickly before the mall closes.

Starting position: Sea Life (Underwater World, pt. marked on map)

The sites you want to visit:

1. The Apple Store
2. Columbia Sportswear
3. Starbucks (1st floor)
4. Ask your friend where they need to go as well.

Planning Goals:

1. Agree upon the quickest path to reach all the stores
2. Estimate the (very) rough time it will take to visit all the places on your list, if you spend 5 minutes at each store.

Show or tell these to the experimenter to complete the task

II. Questionnaire

1. Demographic Questionnaire

Participant ID _______

Date ___/___/___

Demographic Questionnaire

The purpose of this questionnaire is to obtain background information. Your personal identity will not be associated with any of your responses. Your data will be identified only by a unique number and will only be used by the researchers.

Please complete each question by either marking an X or responding in the space provided.

Gender: ☐ Male ☐ Female

Age: ________ years

How long have you lived in the Twin Cities area? ___________________

What is your highest education level completed:

☐ High School / Vocational School

☐ Associates Degree

☐ Bachelor of Arts / Bachelor of Science

☐ Masters

☐ PhD

Are you currently taking any college level classes: ☐ Yes ☐No

What is/was your major field of study?

____

What is your current employment status: ☐ Full Time ☐ Part Time

☐ Retired ☐ Student

☐ Unemployed ☐ Other

If you are currently employed, in what field do you work?

____

How long have you known your partner? Give your best estimate in months or years. If you just met today, write that down.

____

____

How would you rate your familiarity with computers?

Very Comfortable — Extremely Uncomfortable

How often do you use Google Maps, Mapquest, or similar mapping tools?

Every Day — Hardly Ever

How often do you drive or bike in the greater Twin Cities area?

Every Day Hardly Ever

How often do you play computer or video games?

Every Day Hardly Ever

If you have to communicate with someone far away, what tools do you use to collaborate? (Mark all that apply)

☐ Email ☐ Telephon ☐ Fax ☐ Google Docs ☐ Dropbox

☐Video Chat- Skype/Facetime/etc. ☐ Social Media- Facebook/Twitter/etc.

☐ SubVersioN ☐ VNC/Remote Desktop ☐ Others_______________,

2. Post Task Questionnaire

Participant ID _______

condition ________

Post-task Questionnare

Instructions: Think about the collaborativetask you just completed. Please consider the following questions with that task in mind. Place an “X” along each scale at the point that best indicates your experience with the display configuration.

Note!*This form has been designed so that the rating scale is always:*

Good/Easy Bad/Difficult

Mental Demand: How mentally demanding was the task?

Very Low Very High

Physical Demand: How physically demanding was the task?

Very Low (Easy)

Very High (Difficult)

Temporal Demand: How hurried or rushed was the pace of the task?

Very Low (Not Hurried) Very High (Rushed)

Performance: How successful were you in accomplishing the task? (Low score is best)

Very Successful Very Unsuccessful

Effort: How hard did you have to work to accomplish your level of performance?

Very Easy Very Hard

Frustration: How insecure, discouraged, irritated, stressed and annoyed were you?

Not Frustrated Very Frustrated

Solution: To what extent did your solution truly feel like a joint effort?

Very Much a Joint Effort　　　　　　　　　　　　　　Not at All a Joint Effort

Connection: To what degree did you feel disconnected from your teammate?

Strongly Connected　　　　　　　　　　　　　　Strongly Disconnected

Interface: How well do you feel the computer interface allowed you to collaborate seamlessly with your partner?

Seamless Collaboration　　　　　　　　　　　　　　Rough Collaboration

Short Answer: What were most frustrating part(s) about working on this task? For example, your partner, or the computer interface, the difficulty of the task, etc.

____

Cooperation: To what extent were you and your partner a *team* rather than two *separate* individuals?

Team Separate

Interface: To what degree did the computer interface limit your ability to collaborate with your partner?

Not Limiting Greatly Limiting

Solution: To what extent do you feel like you and your partner simply worked your own individual solutions to the problem, and were only able to compare notes afterwards?

Collaborative Solution Individual Solutions

Connection: How much did you feel as if your partner was present with you, while working together on a solution?

Strong Presence Weak Presence

Short Answer: What were the most enjoyable/positive aspect(s) of working on this task?

3. Concluding questionnaire

Participant ID _______

Concluding questionnaire

I felt that all the computer interfaces were roughly the same in terms of usability, enjoyability, and productivity.

Agree	Disagree

Which interface was easiest to use?

Face-to-Face	Phone Call	Drawing Only	Gesture-Voice	Drawing+Gesturing

Which was the most fun?

Face-to-Face	Phone Call	Drawing Only	Gesture-Voice	Drawing+Gesturing

Which would you chooseas a working professional collaborating with people in other locations?

Phone Call	Drawing Only	Gesture-Voice	Drawing+Gesturing

For which interface did you feel most connected to your partner?

Phone Call	Drawing Only	Gesture-Voice	Drawing+Gesturing

Please List the different computer interfaces from 1st favorite to least favorite. For easy reference, the interface names are again:

Face-to-Face	Phone Call	Drawing Only	Gesture-Voice	Drawing+Gesturing

Preference	Interface (write in interface name)
1st - Favorite	
2nd	
3rd	
4th	
5th – Least Favorite	

III. Statistical Analysis

Variables: Face-to-Face
Gesture_Voice
Mark_Gesture_Voice
Mark_Voice
Voice_Only

*. The mean difference is significant at the 0.05 level. Error bars show the 95% confidence interval for each mean.

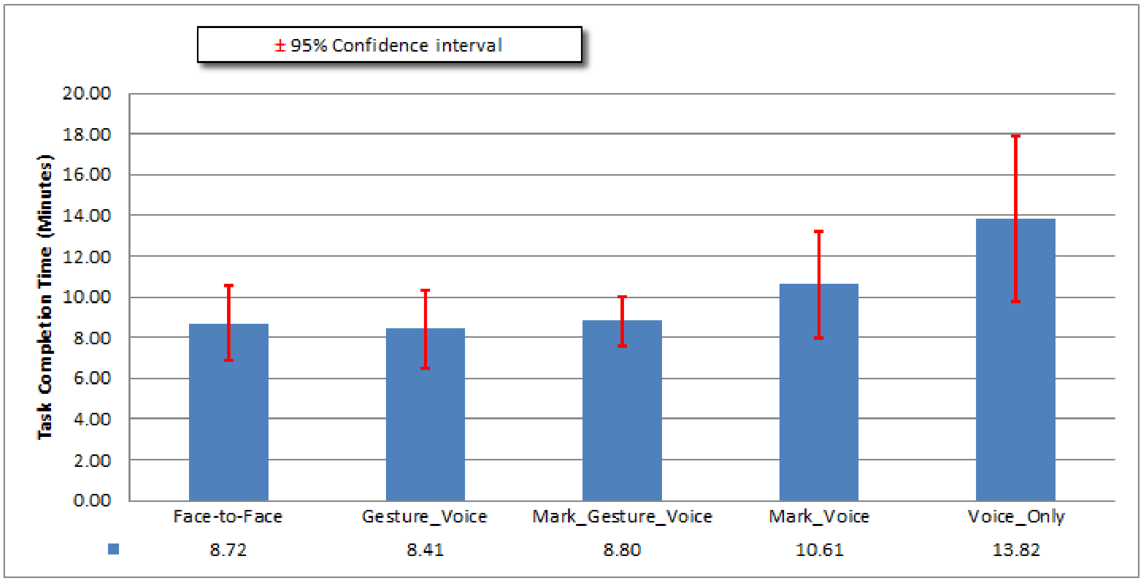
1. Task-completion time in Minutes (significant)

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Variables	411.2197333	4	102.8049333	3.535575844	*0.00981299

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	8.72	1.828534089	0.873632599	3.907003758
Gesture_Voice	20	8.41	1.949875281	0.931606701	4.166271823
Mark_Gesture_Voice	20	8.80	1.180291454	0.563916799	2.521912593
Mark_Voice	20	10.61	2.599959791	1.242202507	5.555298496
Voice_Only	20	13.82	4.067756746	1.94348299	8.691520158



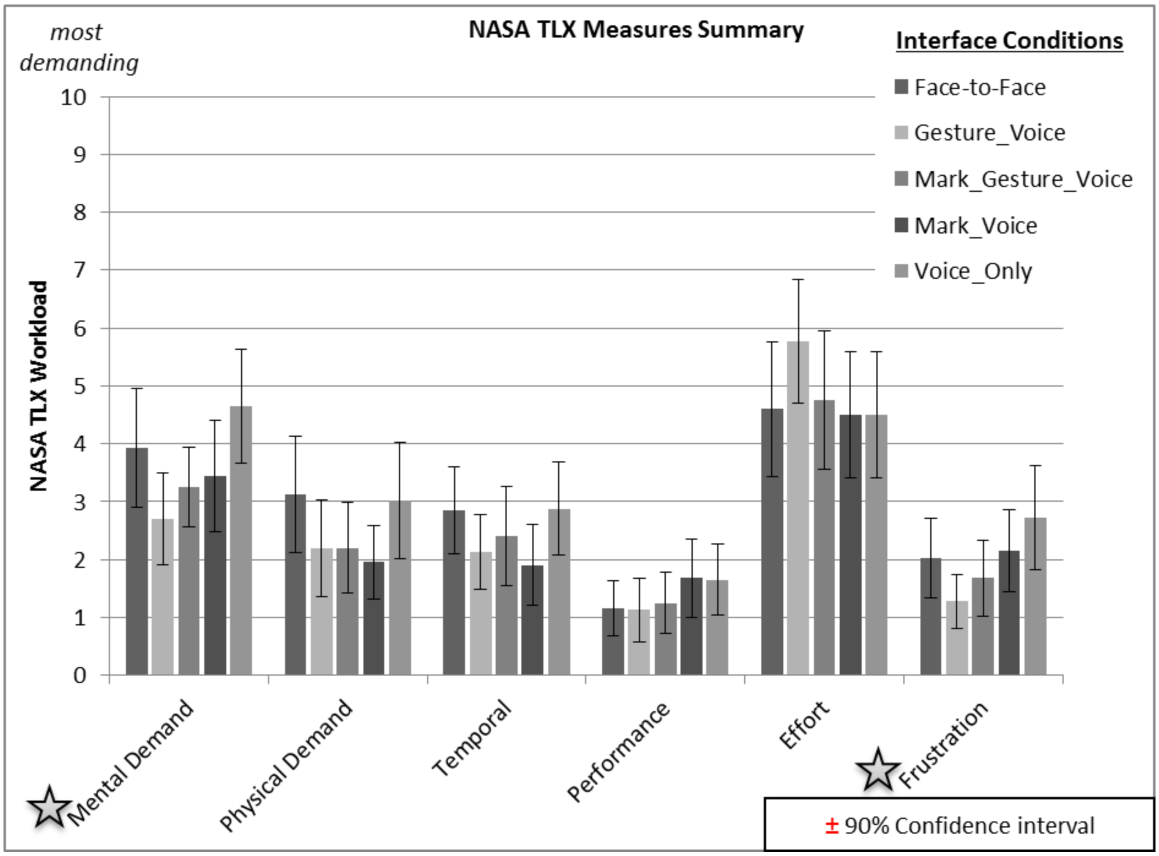
Pairwise Comparison

Task-completion time					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	0.956	0.956	0.059	0.810
Face-to-Face	Mark_Gesture_Voice	0.068	0.068	0.006	0.937
Face-to-Face	Mark_Voice	35.942	35.942	1.558	0.220
Face-to-Face	**Voice_Only**	**260.355**	**260.355**	**5.734**	**** 0.022***
Gesture_Voice	Mark_Gesture_Voice	1.534	1.534	0.129	0.721
Gesture_Voice	Mark_Voice	48.620	48.620	2.017	0.164
Gesture_Voice	**Voice_Only**	**292.861**	**292.861**	**6.305**	**** 0.016***
Mark_Gesture_Voice	Mark_Voice	32.882	32.882	1.767	0.192
Mark_Gesture_Voice	**Voice_Only**	**252.004**	**252.004**	**6.154**	**** 0.018***
Mark_Voice	Voice_Only	102.827	102.827	1.933	0.173

2. Questionnaire

Part 1: NASA TLX (based on a 10-point scale)

NASA TLX Measures Summary Figure: the star symbol indicates the NASA TLX variables where significant pairs were found.



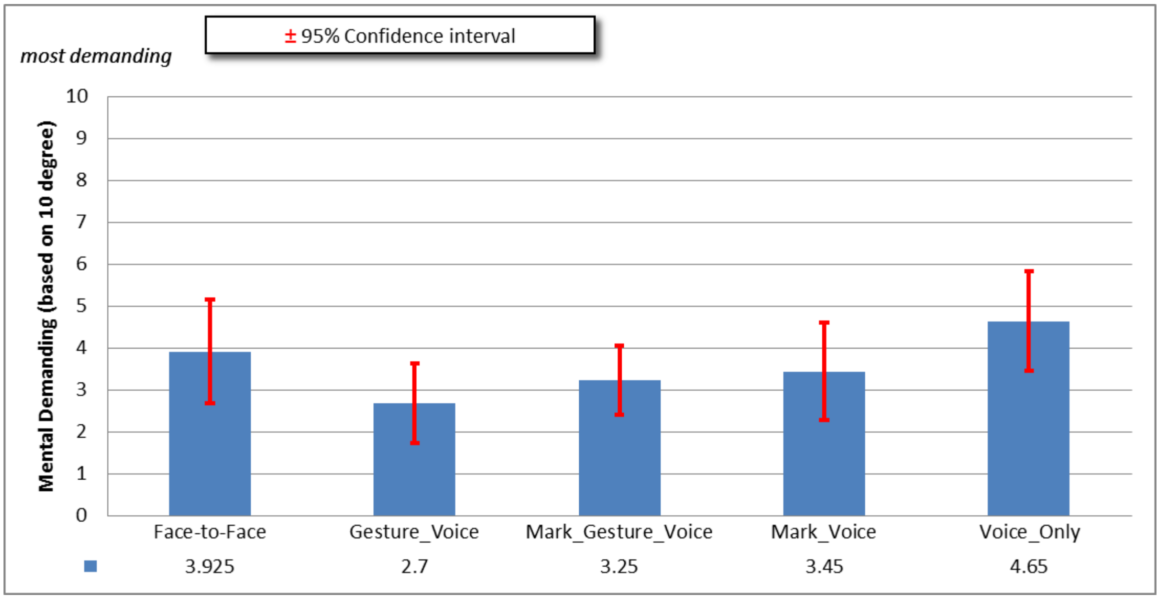
1) Mental Demand (significant)

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Variables	43.26	4	10.815	2.999343162	*0.023556734

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	3.925	1.238889667	0.591913712	2.647118592
Gesture_Voice	20	2.7	0.949479972	0.453640247	2.028740859
Mark_Gesture_Voice	20	3.25	0.826469078	0.394868399	1.765905165
Mark_Voice	20	3.45	1.163614012	0.555948691	2.486278132
Voice_Only	20	4.65	1.186181984	0.566731162	2.534498809



Pairwise Comparison

Mental Demand					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	15.006	15.006	2.698	0.109
Face-to-Face	Mark_Gesture_Voice	4.556	4.556	0.900	0.349
Face-to-Face	Mark_Voice	2.256	2.256	0.342	0.562
Face-to-Face	Voice_Only	5.256	5.256	0.783	0.382
Gesture_Voice	Mark_Gesture_Voice	3.025	3.025	0.836	0.366
Gesture_Voice	Mark_Voice	5.625	5.625	1.093	0.303
Gesture_Voice	**Voice_Only**	**38.025**	**38.025**	**7.216**	**** 0.011***
Mark_Gesture_Voice	Mark_Voice	0.400	0.400	0.086	0.771
Mark_Gesture_Voice	**Voice_Only**	**19.600**	**19.600**	**4.108**	**** 0.050***
Mark_Voice	Voice_Only	14.400	14.400	2.285	0.139

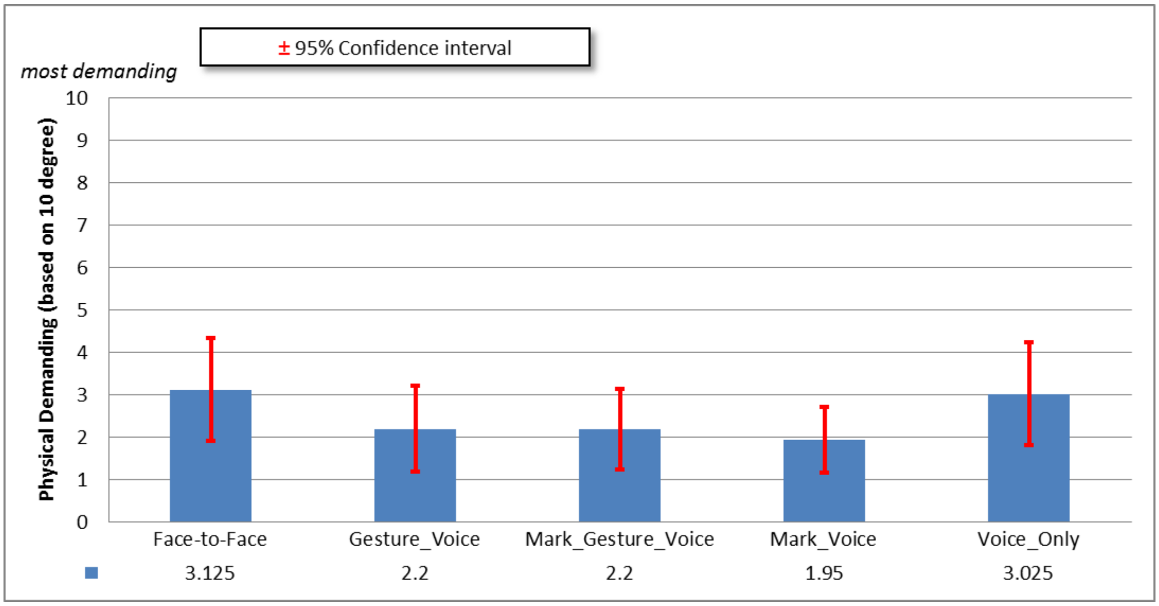
2) Physical Demand

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	91.9	4	22.975	1.1462764	0.3396407

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	3.125	1.209699224	0.577967173	2.584747774
Gesture_Voice	20	2.2	1.014062818	0.484496483	2.166734142
Mark_Gesture_Voice	20	2.2	0.952510557	0.455088193	2.035216269
Mark_Voice	20	1.95	0.777598077	0.371518931	1.661483167
Voice_Only	20	3.025	1.218245925	0.582050595	2.603009392



Pairwise Comparison

Physical Demand					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	34.225	34.225	1.504	0.228
Face-to-Face	Mark_Gesture_Voice	34.225	34.225	1.581	0.216
Face-to-Face	Mark_Voice	55.225	55.225	2.925	0.095
Face-to-Face	Voice_Only	0.4	0.4	0.015	0.904
Gesture_Voice	Mark_Gesture_Voice	0	0	0.000	1.000
Gesture_Voice	Mark_Voice	2.5	2.5	0.168	0.684
Gesture_Voice	Voice_Only	27.225	27.225	1.187	0.283
Mark_Gesture_Voice	Mark_Voice	2.5	2.5	0.181	0.673
Mark_Gesture_Voice	Voice_Only	27.225	27.225	1.247	0.271
Mark_Voice	Voice_Only	46.225	46.225	2.424	0.128

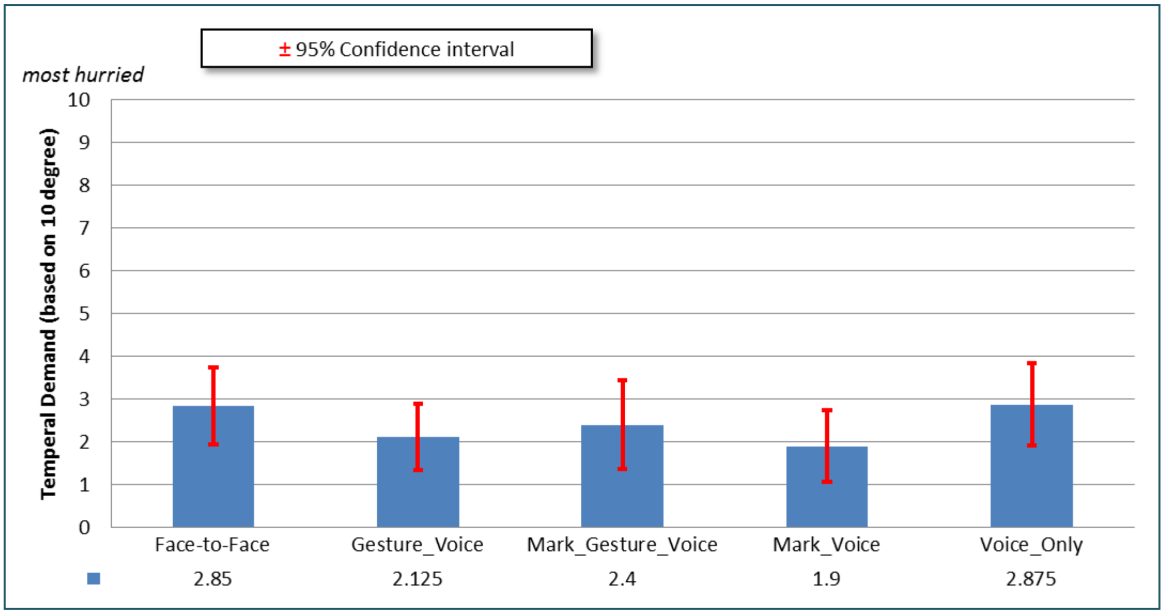
3) Temporal Demand

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	14.985	4	3.74625	0.98715415	0.418442815

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	2.85	0.901843004	0.430880372	1.926955603
Gesture_Voice	20	2.125	0.773789714	0.36969938	1.65334589
Mark_Gesture_Voice	20	2.4	1.042649353	0.498154489	2.227814601
Mark_Voice	20	1.9	0.84747414	0.404904147	1.810786395
Voice_Only	20	2.875	0.96893501	0.462935428	2.070310173



Pairwise Comparison

Temporal Demand					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	5.256	5.256	1.631	0.209
Face-to-Face	Mark_Gesture_Voice	2.025	2.025	0.467	0.499
Face-to-Face	Mark_Voice	9.025	9.025	2.581	0.116
Face-to-Face	Voice_Only	0.006	0.006	0.002	0.969
Gesture_Voice	Mark_Gesture_Voice	0.756	0.756	0.197	0.660
Gesture_Voice	Mark_Voice	0.506	0.506	0.168	0.684
Gesture_Voice	Voice_Only	5.625	5.625	1.603	0.213
Mark_Gesture_Voice	Mark_Voice	2.500	2.500	0.607	0.441
Mark_Gesture_Voice	Voice_Only	2.256	2.256	0.488	0.489
Mark_Voice	Voice_Only	9.506	9.506	2.513	0.121

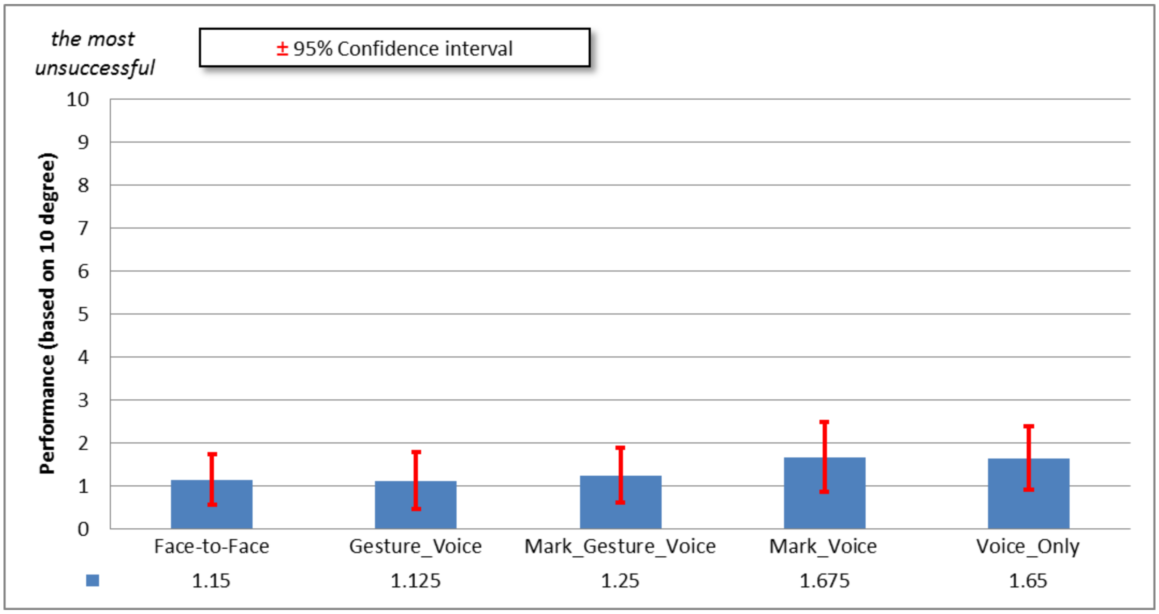
4) Performance

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	5.885	4	1.47125	0.670594938	0.613974992

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	1.15	0.578702689	0.276491172	1.236506113
Gesture_Voice	20	1.125	0.66132781	0.315967603	1.413050079
Mark_Gesture_Voice	20	1.25	0.646452229	0.308860384	1.381265629
Mark_Voice	20	1.675	0.818673265	0.391143735	1.749247959
Voice_Only	20	1.65	0.736482184	0.351874679	1.573631402



Pairwise Comparison

Performance					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	0.006	0.006	0.004	0.953
Face-to-Face	Mark_Gesture_Voice	0.100	0.100	0.058	0.811
Face-to-Face	Mark_Voice	2.756	2.756	1.201	0.280
Face-to-Face	Voice_Only	2.500	2.500	1.248	0.271
Gesture_Voice	Mark_Gesture_Voice	0.156	0.156	0.080	0.779
Gesture_Voice	Mark_Voice	3.025	3.025	1.196	0.281
Gesture_Voice	Voice_Only	2.756	2.756	1.232	0.274
Mark_Gesture_Voice	Mark_Voice	1.806	1.806	0.727	0.399
Mark_Gesture_Voice	Voice_Only	1.600	1.600	0.730	0.398
Mark_Voice	Voice_Only	0.006	0.006	0.002	0.962

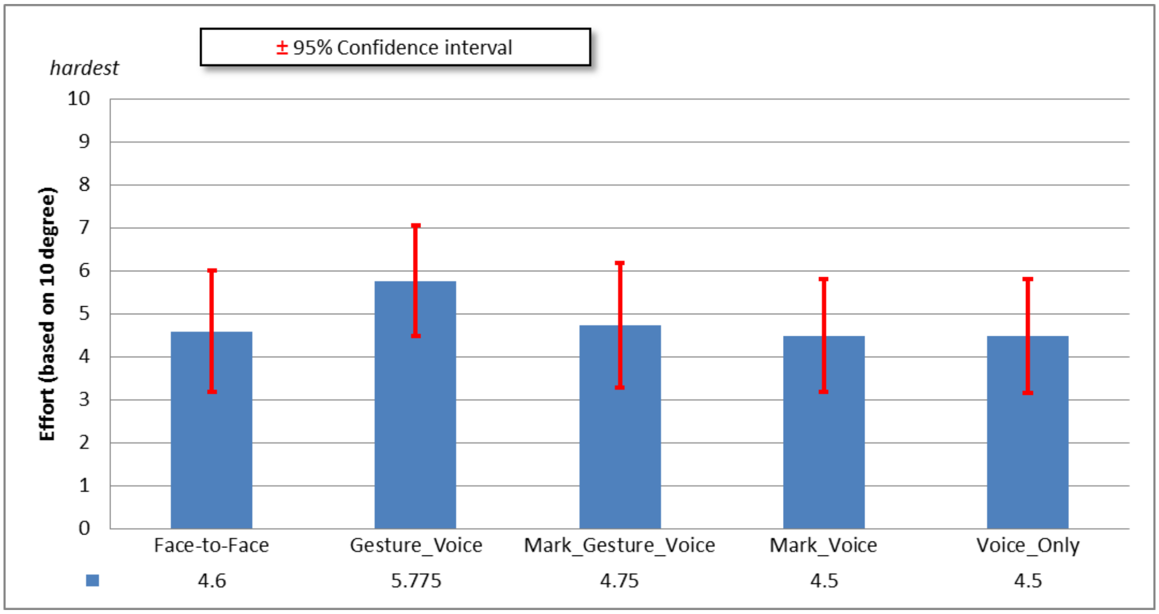
5) Effort

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	25.02472754	4	6.256181885	0.743526579	0.564721614

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	4.6	1.40527433	0.671408574	3.002630426
Gesture_Voice	20	5.775	1.28950053	0.616094448	2.755258131
Mark_Gesture_Voice	20	4.75	1.44949326	0.692535388	3.097112407
Mark_Voice	20	4.5	1.312813482	0.627232854	2.805070597
Voice_Only	20	4.5	1.325920098	0.633494901	2.833075324



Pairwise Comparison

Effort					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	17.174	17.174	2.072	0.158
Face-to-Face	Mark_Gesture_Voice	0.892	0.892	0.096	0.759
Face-to-Face	Mark_Voice	0.027	0.027	0.003	0.955
Face-to-Face	Voice_Only	0.027	0.027	0.003	0.955
Gesture_Voice	Mark_Gesture_Voice	10.506	10.506	1.223	0.276
Gesture_Voice	Mark_Voice	16.256	16.256	2.103	0.155
Gesture_Voice	Voice_Only	16.256	16.256	2.082	0.157
Mark_Gesture_Voice	Mark_Voice	0.625	0.625	0.072	0.790
Mark_Gesture_Voice	Voice_Only	0.625	0.625	0.071	0.791
Mark_Voice	Voice_Only	0	0	1.02	0.97

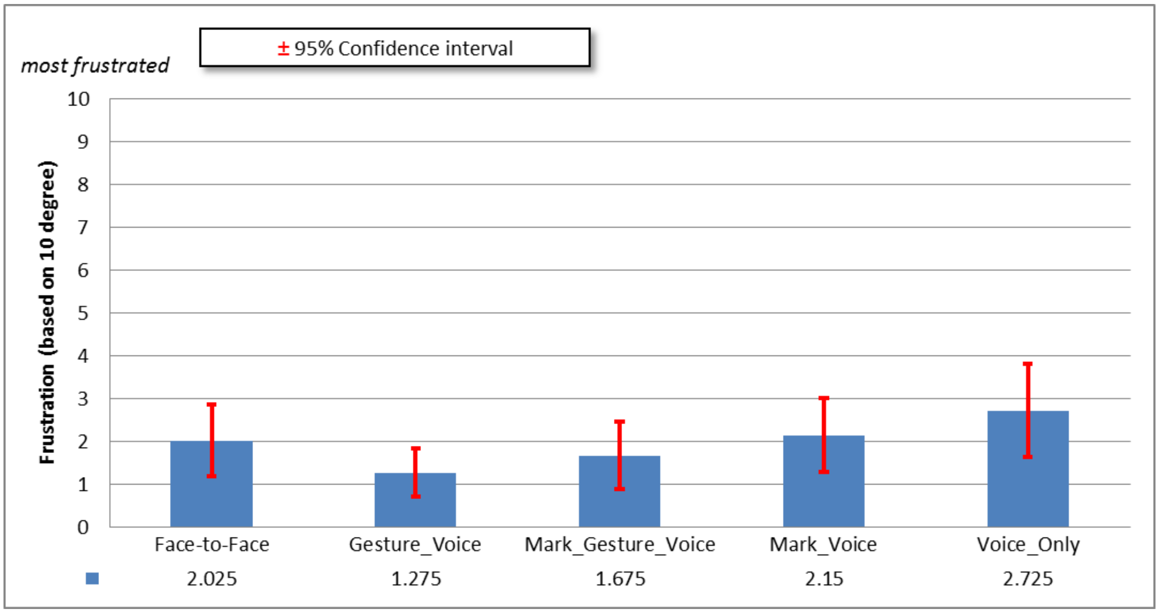
6) Frustration (significant)

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	93.51	19	4.921578947	1.720594324	* 0.04806938

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	2.025	0.840216407	0.401436565	1.795278896
Gesture_Voice	20	1.275	0.555191528	0.265258067	1.186270139
Mark_Gesture_Voice	20	1.675	0.797271	0.380918212	1.703518032
Mark_Voice	20	2.15	0.865976451	0.413744137	1.850320029
Voice_Only	20	2.725	1.088295738	0.519963308	2.325346607



Pairwise Comparison

Frustration					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	5.625	5.625	2.430	0.127
Face-to-Face	Mark_Gesture_Voice	1.225	1.225	0.400	0.531
Face-to-Face	Mark_Voice	0.156	0.156	0.047	0.829
Face-to-Face	Voice_Only	4.900	4.900	1.136	0.293
Gesture_Voice	Mark_Gesture_Voice	1.600	1.600	0.743	0.394
Gesture_Voice	Mark_Voice	7.656	7.656	3.170	0.083
Gesture_Voice	**Voice_Only**	**21.025**	**21.025**	**6.171**	**** 0.018***
Mark_Gesture_Voice	Mark_Voice	2.256	2.256	0.713	0.404
Mark_Gesture_Voice	Voice_Only	11.025	11.025	2.654	0.112
Mark_Voice	Voice_Only	3.306	3.306	0.749	0.392

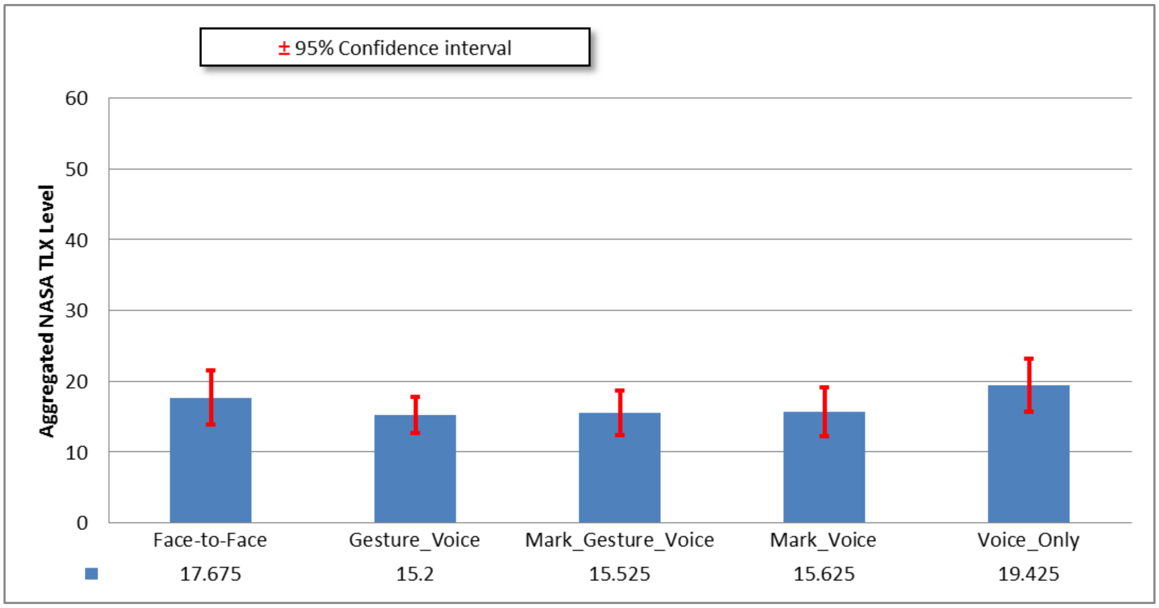
7) Aggregated NASA TLX

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	263.24	4	65.81	1.252607115	0.294170365

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	17.675	3.858057351	1.84329332	8.243458331
Gesture_Voice	20	15.2	2.557116666	1.221733023	5.463756179
Mark_Gesture_Voice	20	15.525	3.17193479	1.515479341	6.777429648
Mark_Voice	20	15.625	3.454970762	1.650707584	7.382188735
Voice_Only	20	19.425	3.757389425	1.795196441	8.028362552



Pairwise Comparison

Aggregated NASA TLX					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	61.25625	61.25625	1.253	0.270
Face-to-Face	Mark_Gesture_Voice	46.225	46.225	0.812	0.373
Face-to-Face	Mark_Voice	42.025	42.025	0.686	0.413
Face-to-Face	Voice_Only	30.625	30.625	0.463	0.501
Gesture_Voice	Mark_Gesture_Voice	1.05625	1.05625	0.028	0.868
Gesture_Voice	Mark_Voice	1.80625	1.80625	0.043	0.837
Gesture_Voice	Voice_Only	178.50625	178.50625	3.786	0.059
Mark_Gesture_Voice	Mark_Voice	0.1	0.1	0.002	0.965
Mark_Gesture_Voice	Voice_Only	152.1	152.1	2.756	0.105
Mark_Voice	Voice_Only	144.4	144.4	2.428	0.127

Part 2: Short Questions

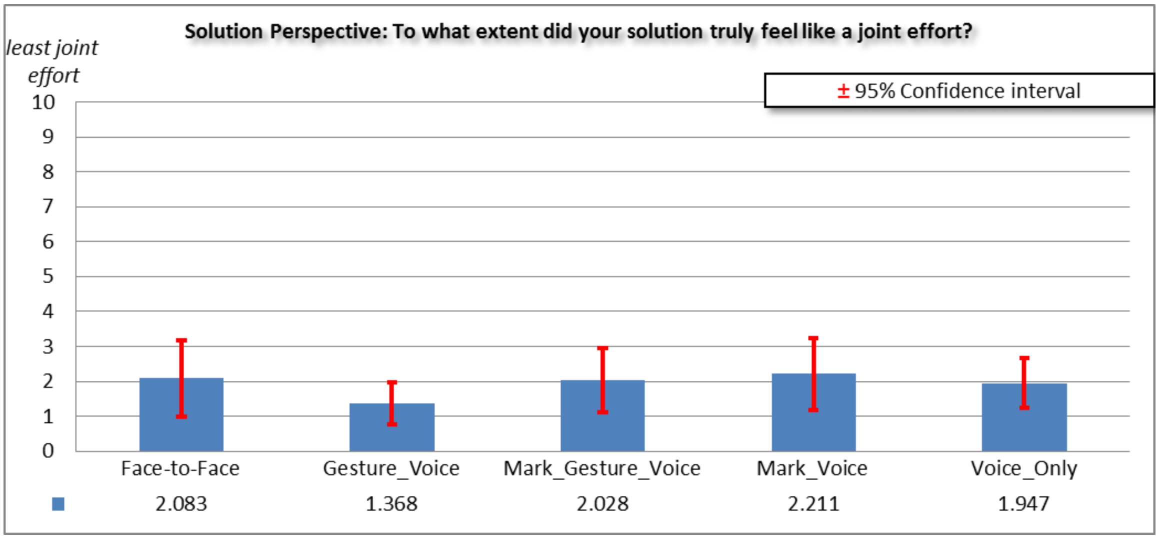
1) Solution Perspective: To what extent did your solution truly feel like a joint effort?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	30.74	4	7.685	0.59856932	0.664559486

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	18	2.083333333	1.086328432	0.51489259	2.184504251
Gesture_Voice	19	1.368421053	0.605643401	0.28827503	1.256561725
Mark_Gesture_Voice	18	2.027777778	0.912470538	0.432488283	1.834892385
Mark_Voice	19	2.210526316	1.020282298	0.485635457	2.116835882
Voice_Only	19	1.947368421	0.713519635	0.339622117	1.480378488



Pairwise Comparison

Solution Perspective: To what extent did your solution truly feel like a joint effort?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	18.225	18.225	1.500	0.228
Face-to-Face	Mark_Gesture_Voice	0.1	0.1	0.006	0.936
Face-to-Face	Mark_Voice	0.625	0.625	0.035	0.853
Face-to-Face	Voice_Only	0.625	0.625	0.046	0.831
Gesture_Voice	Mark_Gesture_Voice	15.625	15.625	1.625	0.210
Gesture_Voice	Mark_Voice	25.6	25.6	2.106	0.155
Gesture_Voice	Voice_Only	12.1	12.1	1.569	0.218
Mark_Gesture_Voice	Mark_Voice	1.225	1.225	0.079	0.780
Mark_Gesture_Voice	Voice_Only	0.225	0.225	0.021	0.887
Mark_Voice	Voice_Only	2.5	2.5	0.185	0.670

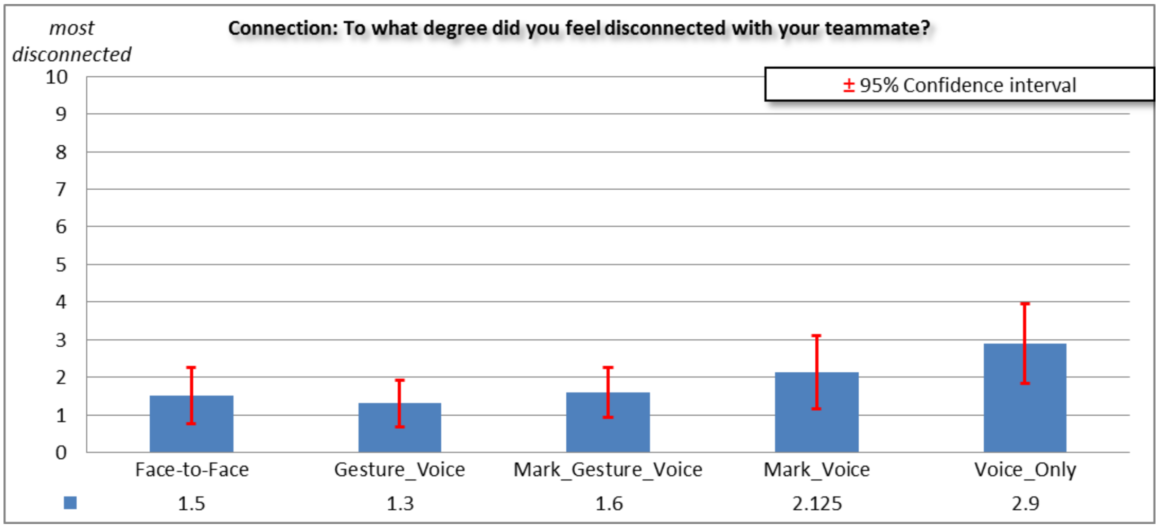
2) Connection: To what degree did you feel disconnected with your teammate (significant)?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	33.19	4	8.2975	2.629842779	*0.0390977

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	1.5	0.747744627	0.357255621	1.597695709
Gesture_Voice	20	1.3	0.627907231	0.3	1.341640786
Mark_Gesture_Voice	20	1.6	0.664479855	0.31747358	1.41978501
Mark_Voice	20	2.125	0.971904933	0.464354391	2.076655966
Voice_Only	20	2.9	1.056379854	0.504714615	2.257152375



Pairwise Comparison

Connection: To what degree did you feel disconnected with your teammate?					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	0.400	0.400	0.184	0.671
Face-to-Face	Mark_Gesture_Voice	0.100	0.100	0.044	0.835
Face-to-Face	Mark_Voice	3.906	3.906	1.138	0.293
Face-to-Face	**Voice_Only**	**19.600**	**19.600**	**5.126**	**** 0.029***
Gesture_Voice	Mark_Gesture_Voice	0.900	0.900	0.472	0.496
Gesture_Voice	Mark_Voice	6.806	6.806	2.227	0.144
Gesture_Voice	**Voice_Only**	**25.600**	**25.600**	**7.426**	**** 0.010***
Mark_Gesture_Voice	Mark_Voice	2.756	2.756	0.871	0.357
Mark_Gesture_Voice	**Voice_Only**	**16.900**	**16.900**	**4.754**	**** 0.036***
Mark_Voice	Voice_Only	6.006	6.006	1.277	0.266

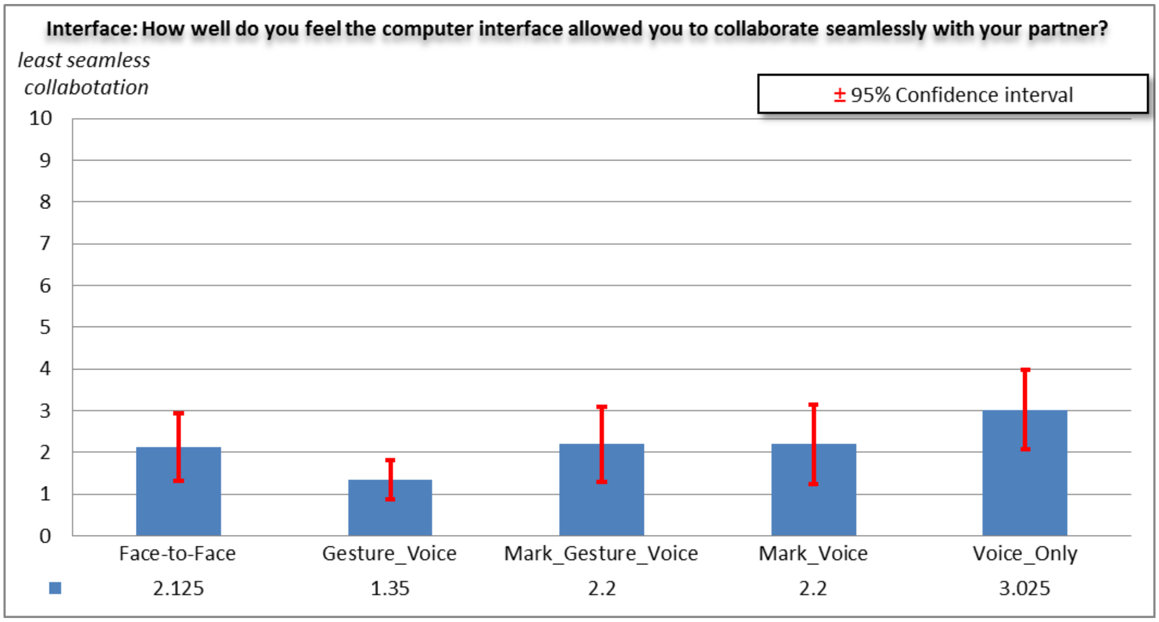
3) Interface: How well do you feel the computer Interface allows you to collaborate seamlessly with your partner (significant)?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Variables	28.135	4	7.03375	2.553268216	*0.045669555

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	2.125	0.817263884	0.390470364	1.746236555
Gesture_Voice	20	1.35	0.474739986	0.226820123	1.01437043
Mark_Gesture_Voice	20	2.2	0.896393752	0.428276841	1.91531226
Mark_Voice	20	2.2	0.949479972	0.453640247	2.028740859
Voice_Only	20	3.025	0.949707601	0.453749003	2.029227232



Pairwise Comparison

Interface: How well do you feel the computer Interface allows you to collaborate seamlessly with your partner?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	6.006	6.006	2.945	0.094
Face-to-Face	Mark_Gesture_Voice	0.056	0.056	0.017	0.898
Face-to-Face	Mark_Voice	0.056	0.056	0.016	0.901
Face-to-Face	Voice_Only	8.100	8.100	2.260	0.141
Gesture_Voice	Mark_Gesture_Voice	7.225	7.225	3.076	0.088
Gesture_Voice	Mark_Voice	7.225	7.225	2.809	0.102
Gesture_Voice	**Voice_Only**	**28.056**	**28.056**	**10.903**	**** 0.002***
Mark_Gesture_Voice	Mark_Voice	0.0001	0.0001	0.001	0.999
Mark_Gesture_Voice	Voice_Only	6.806	6.806	1.748	0.194
Mark_Voice	Voice_Only	6.806	6.806	1.653	0.206

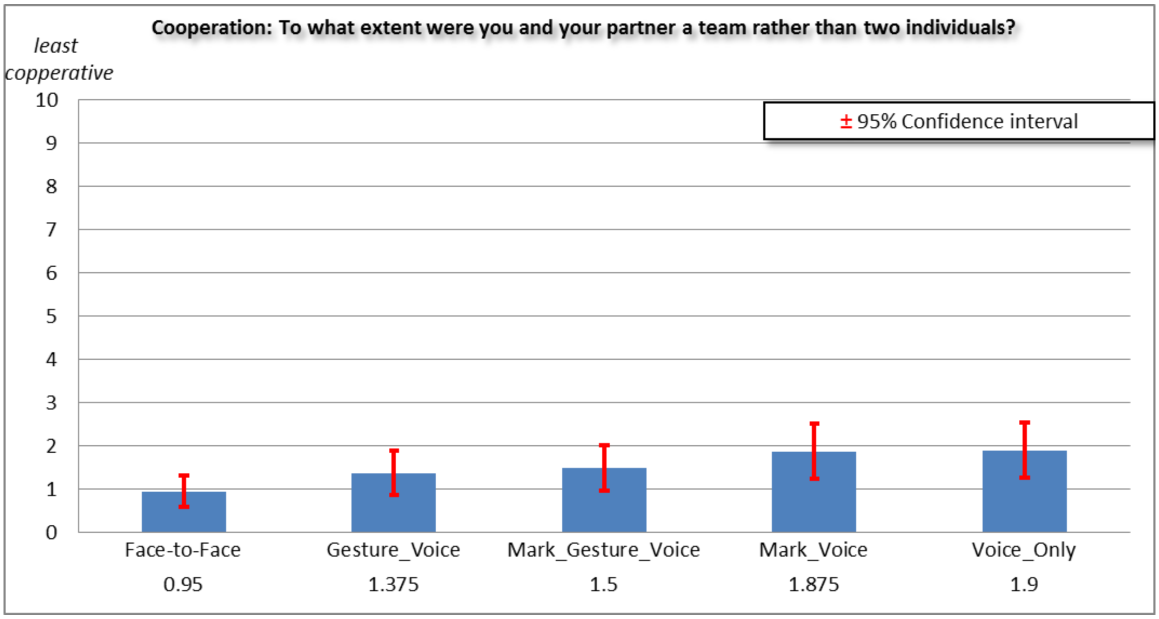
4) Cooperation: To what extent were you and your partner a team rather than two individuals?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	12.335	4	3.08375	2.268780252	0.06745522

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	0.95	0.37116446	0.177334059	0.793062021
Gesture_Voice	20	1.375	0.519802153	0.248349817	1.110654146
Mark_Gesture_Voice	20	1.5	0.526002796	0.251312345	1.123902974
Mark_Voice	20	1.875	0.639166449	0.305379402	1.365698202
Voice_Only	20	1.9	0.628824554	0.300438276	1.343600818



Pairwise Comparison

Cooperation: To what extent were you and your partner a team rather than two individuals?					
Pairs		Sum of Squares	Mean Square	F	P
Face-to-Face	Gesture_Voice	1.806	1.806	1.940	0.172
Face-to-Face	Mark_Gesture_Voice	3.025	3.025	3.197	0.082
Face-to-Face	Mark_Voice	8.556	8.556	6.861	0.068
Face-to-Face	Voice_Only	9.025	9.025	7.415	0.289
Gesture_Voice	Mark_Gesture_Voice	0.156	0.156	0.125	0.725
Gesture_Voice	Mark_Voice	2.500	2.500	1.614	0.212
Gesture_Voice	Voice_Only	2.756	2.756	1.814	0.186
Mark_Gesture_Voice	Mark_Voice	1.406	1.406	0.899	0.349
Mark_Gesture_Voice	Voice_Only	1.600	1.600	1.043	0.314
Mark_Voice	Voice_Only	0.006	0.006	0.003	0.954

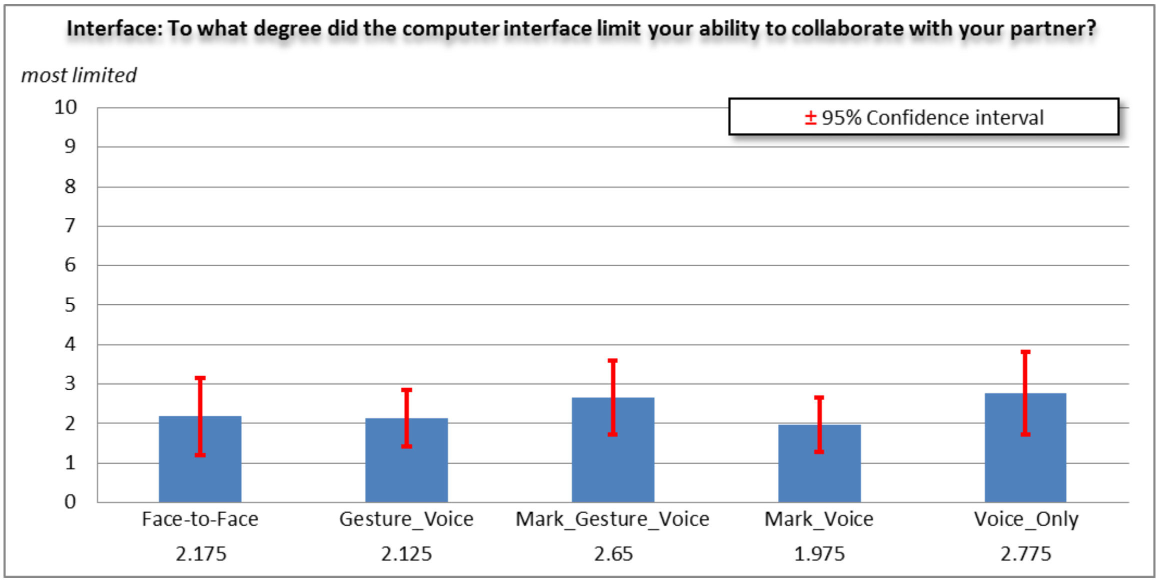
5) Interface: To what degree did the computer interface limit your ability to collaborate with your partner?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	37.15478469	4	9.288696172	0.652334297	0.62665456

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	2.175	0.97012407	0.463503534	2.072850821
Gesture_Voice	20	2.125	0.719758638	0.343884543	1.537898429
Mark_Gesture_Voice	20	2.65	0.942472644	0.450292303	2.013768397
Mark_Voice	20	1.975	0.693658397	0.331414433	1.482130401
Voice_Only	20	2.775	1.042303773	0.497989378	2.227076205



Pairwise Comparison

Interface: To what degree did the computer interface limit your ability to collaborate with your partner?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	1.054	1.054	0.080	0.779
Face-to-Face	Mark_Gesture_Voice	5.066	5.066	0.305	0.584
Face-to-Face	Mark_Voice	3.854	3.854	0.301	0.586
Face-to-Face	Voice_Only	9.188	9.188	0.497	0.485
Gesture_Voice	Mark_Gesture_Voice	11.025	11.025	0.859	0.360
Gesture_Voice	Mark_Voice	0.9	0.9	0.099	0.755
Gesture_Voice	Voice_Only	16.9	16.9	1.154	0.290
Mark_Gesture_Voice	Mark_Voice	18.225	18.225	1.458	0.235
Mark_Gesture_Voice	Voice_Only	0.625	0.625	0.035	0.853
Mark_Voice	Voice_Only	25.6	25.6	1.789	0.189

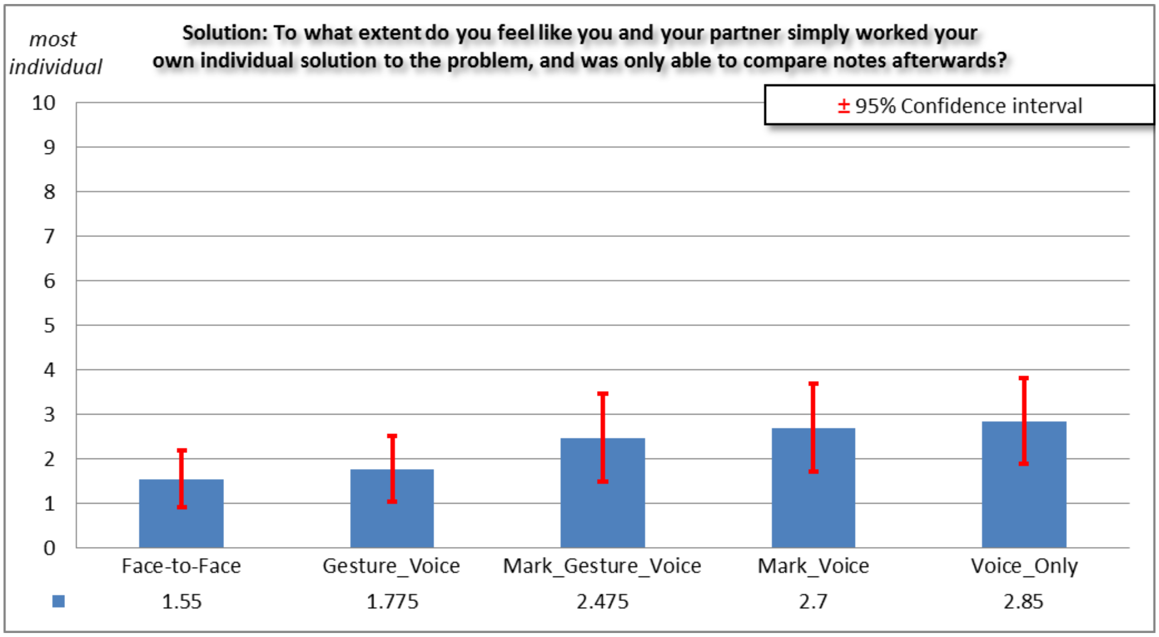
6) Solution: To what extent do you feel like you and your partner simply worked your own individual solution to the problem, and was only able to compare notes afterwards (Significant in pairwise comparisons)?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Groups	31.8	4	7.95	2.321	0.0623

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	1.55	0.639279167	0.305433256	1.365939045
Gesture_Voice	20	1.775	0.665671553	0.318042946	1.422331296
Mark_Gesture_Voice	20	2.475	0.979584593	0.46802356	2.09306499
Mark_Voice	20	2.7	0.993969007	0.474896111	2.123799971
Voice_Only	20	2.85	0.960645383	0.45897483	2.05259784



Pairwise Comparison

Solution: To what extent do you feel like you and your partner simply worked your own individual solution to the problem, and was only able to compare notes afterwards?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	0.506	0.506	0.236	0.630
Face-to-Face	Mark_Gesture_Voice	8.556	8.556	2.739	0.106
Face-to-Face	**Mark_Voice**	**13.225**	**13.225**	**4.148**	****0.049***
Face-to-Face	**Voice_Only**	**16.9**	**16.9**	**5.560**	****0.024***
Gesture_Voice	Mark_Gesture_Voice	4.900	4.900	1.438	0.238
Gesture_Voice	Mark_Voice	8.556	8.556	2.464	0.125
Gesture_Voice	Voice_Only	11.556	11.556	3.477	0.070
Mark_Gesture_Voice	Mark_Voice	0.506	0.506	0.114	0.738
Mark_Gesture_Voice	Voice_Only	1.406	1.406	0.327	0.571
Mark_Voice	Voice_Only	0.225	0.225	0.052	0.822

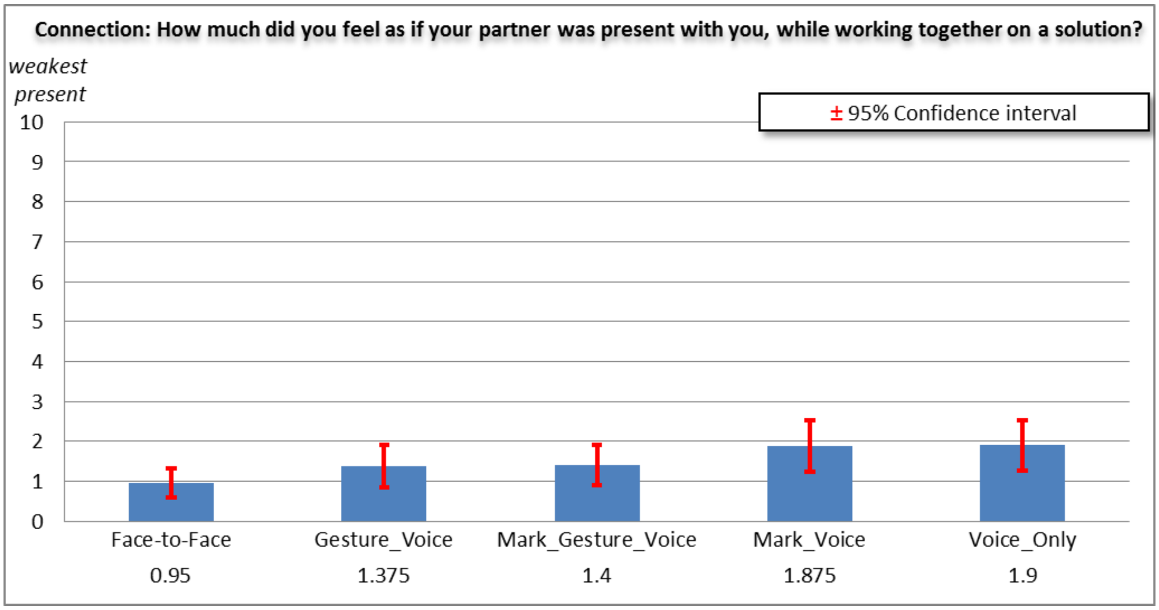
7) Connection: How much did you feel as if your partner was present with you, while working together on a solution (significant)?

ANOVA Analysis

	Sum of Squares	Degrees of Freedom	Mean Square	F	P
Between Variables	12.335	4	3.08375	4.225457496	*0.003840242

Mean Analysis

	N	Mean	95% Conf. (±)	Std.Error	Std.Dev.
Face-to-Face	20	0.95	0.371	0.177	0.793
Gesture_Voice	20	1.375	0.520	0.248	1.111
Mark_Gesture_Voice	20	1.4	0.507	0.242	1.083
Mark_Voice	20	1.875	0.639	0.305	1.366
Voice_Only	20	1.9	0.629	0.300	1.344



Pairwise Comparison

Connection: How much did you feel as if your partner was present with you, while working together on a solution?					
Pairs		**Sum of Squares**	**Mean Square**	**F**	**P**
Face-to-Face	Gesture_Voice	7.225	7.225	1.940	0.172
Face-to-Face	Mark_Gesture_Voice	8.1	8.1	2.247	0.142
Face-to-Face	**Mark_Voice**	**34.225**	**34.225**	**6.861**	****0.013***
Face-to-Face	**Voice_Only**	**36.1**	**36.1**	**7.415**	****0.010***
Gesture_Voice	Mark_Gesture_Voice	0.025	0.025	0.005	0.943
Gesture_Voice	Mark_Voice	10	10	1.614	0.212
Gesture_Voice	Voice_Only	11.025	11.025	1.814	0.186
Mark_Gesture_Voice	Mark_Voice	9.025	9.025	1.485	0.231
Mark_Gesture_Voice	Voice_Only	10	10	1.678	0.203
Mark_Voice	Voice_Only	0.025	0.025	0.003	0.954