# Task Conflict and Language Differences: Opportunities for Videoconferencing?

Gayna Williams
Microsoft Corporation, USA
*GaynaW@Microsoft.com*

**Abstract:** Considerable research has found that adding audio to desktop conferencing improves problem-solving, but that there is no additional benefit from adding video This paper describes an experiment that supports these earlier findings, while suggesting that video may provide significant benefit when used in tasks involving speakers with different priorities and different linguistic capabilities The level of conflict in the experimental setting is consistent with that found in work settings, but is higher than that in most experimental situations Tasks involving a mix of native and non-native English speakers are not universal but may be increasingly common Thus, these findings provide encouragement to those working to improve videoconferencing technology

## Introduction

Studies contrasting audio, audio with video, and face to face problem-solving date back at least to Chapanis, Ochsman, Parrish, & Weeks (1972), who found no benefit in adding high-quality video to audio. In reviewing the history of research into the use of video to support communication and distributed work, Egido (1988) wrote "videoconferencing has been commercially available for over two decades, and despite consistently brilliant market forecasts, to date it has failed except in limited niche markets... Results (of systematic research) generally point to the dubious value of adding a visual channel... performance does not improve



*J. Hughes et al. (eds.), Proceedings of the Fifth European Conference on Computer Supported Cooperative Work,* 97-108

significantly... However, it can provide a sense of social presence and mutual knowledge (and) add to its desirability or appeal..."

The years since the first studies have seen tremendous technological innovation, ingenious attacks on identified weaknesses in video systems, attention focused on behavioral and social considerations, and much-improved interfaces. Despite this work and dramatic price-performance improvements, desktop videoconferencing has yet to spread appreciably. Studies have continued to find marginal or non-existent advantages in adding video to audio in supporting distributed groups. Examples include Sellen (1992), and Olson, Olson & Meader (1995), whose very careful study concluded "the average judged quality of the groups supported by audio plus video was *not* significantly higher than that for groups supported by audio only," and Doerry (1996), whose dissertation concluded that what he had found "bodes ill for the current crop of technologically-mediated environments." In a recent review, Whittaker (1995) echoed Egido's earlier remarks: "Evaluations of videoconferencing systems show that previous work has overestimated the importance of video at the expense of audio." This has been true even though many experimental systems have been of high quality (including computer controlled analog, which eliminates bandwidth constraints).

These reports and others do, however, include grounds for optimism. Video supports a sense of presence and mutual knowledge, which may be motivating, as reported by Egido (1988, quoted above) and Olson, Olson and Meader (1995) among others. Video can be useful for tasks in which the focus of the task (and the video) is on large objects (e.g., Magee, 1992). In addition, several research efforts have looked at video to provide opportunities for unplanned, informal contact (in contrast to substituting video for task-centered meetings), (e.g., Root, 1988; Fish et al., 1990; Olson & Bly, 1991; Dourish & Bly, 1992; Buxton, 1992, Mantei et al., 1991; Tang & Rua, 1994).

The experiment described below does not address this wide range of possible video use. This study focuses on the use of video to support distributed workgroups carrying out assigned tasks in concert, which remains a major concern. It was motivated as part of a RACE (Research and development on Advanced Communications in Europe) project to support concurrent engineering within the European automotive industry.

Support for multinational efforts brings distributed work to the fore and raises unexplored issues. These issues are particularly pertinent considering the development of computer networked global communities, which are leading to an increase in international business partnerships. The participation by non-native speakers of the language being used could affect the use and usefulness of video. With reduced linguistic competence, nonverbal cues to meaning or intent may be sought. Nonverbal communication skills are culturally based (Argyle, 1969), so we cannot assume they will be useful in all such situations. However, cues may be more readily understood across cultures with relatively high levels of contact,

such as among Europeans. This study, although carried out in a laboratory, contrasts groups of native English speakers with groups of Europeans whose first language is not English (i.e., non-native English speakers), and groups that mix native English speakers with non-native speakers. (Because this work was motivated by research into the communication requirements of Europeans in the automotive industry, participants from countries outside Europe were excluded from this particular study.)

Another issue addressed in this study is that of goal conflict or differing priorities within workgroups. In many laboratory situations, participants may generate and then promote or defend their ideas, but have less stake in the outcome than in many work situations. Laboratory studies of non-contentious tasks have generally shown no benefit of adding video to audio in problem solving (e.g., Weeks & Chapanis, 1976; Olson, Olson & Meader, 1995; Doerry, 1996). Contention in work pairs was investigated by Weeks and Chapanis (1976), who found no differences in performance between audio and video communication media. However, contention could direct participants to nonverbal cues in situations involving more than two people, where group dynamics and social cues may be more variable. In my within-subjects study, a non-contentious design task is followed by a task in which group members were randomly assigned the roles of ergonomist (human factors engineer), accountant (with fiscal responsibility), and designer. The resulting level of contention, although typical of development projects, did affect the use of video, especially in the presence of language differences. The experiment, results and discussion are reported in the following sections.

# Experimental Method

## Design

Table I identifies the three factors that were varied in the experiment. All tasks were completed by groups of 3 participants working in separate rooms with earphones and desktop conferencing software (Timbuktu).

| | |
|---|---|
| Teleconferencing conditions | 1. Audio (A)<br>2. Audio and video (AV) |
| Design task conditions | 1. Non-contentious (NC)<br>2. Contentious (C) |
| Language conditions | 1. Native speakers (NS)<br>2. Non-native speakers (NNS)<br>3. Mixed speakers (MIX) |

Table 1. Experimental conditions

1. In half of the tasks, participants were linked only by audio. In the other half, participants also had a video link, consisting of two monitors to the right of the computer display. The order of the communication media was counterbalanced across the groups.
2. As detailed below, each group carried out one non-contentious and one contentious design task. Each individual then completed a questionnaire.
3. Each group consisted of 3 subjects. One-third of the groups consisted of native English speakers (NS), one-third consisted of non-native English speakers (NNS), and one-third consisted of one native and two non-native English speakers. No group had non-native speakers from the same country. Group members had not previously worked together.

## Subjects

All subjects in the experiment were students at Loughborough University of Technology. Each was paid £7.50 for participating in the two-hour experiment. Sixteen were native English speakers (8 male, 8 female); the other 20 were from seven different European countries (11 female, 9 male). (Gender was noted but not a factor in the study.)

## Apparatus

Each participant worked in a room with a Macintosh computer running MacDraw and Timbuktu software, a headphone for 3-way audio communication, a Sony CCDGIE video camera, and two video monitors for video support in one of the two tasks. In addition, a camera was focused on the video monitors to record the actions of each participant, and audio communication was recorded.

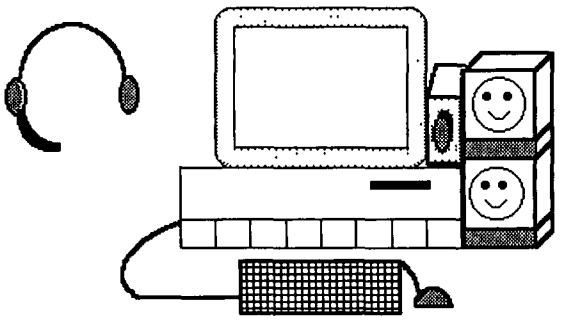

Figure 1. A diagram of the apparatus used in the study. Each room contained the same setup

Analog video was displayed on video monitors to the right of the computer display (Figure 1). This approach, used in CRUISER (Root, 1988) and other systems, creates a "seam" but allows the number and duration of glances at the video monitors to be determined from the video recordings of the participants.

## Task

The task was based on a design exercise devised by Brian Shackel to train people in skills such as negotiation, compromise, and team work, all skills which are required in industrial design settings. Participants in this particular exercise are given the task of individually designing a control panel for a stereo sound system from a set of component requirements. After completing this task the students are required to design the control panel in groups; this can lead to discussion, disagreement, and compromise of their individual designs in order to achieve a group design.

The task was modified to accommodate the time constraints of the study session and the skill set of the subject population (i.e., they were not trained designers and perhaps wouldn't be as sufficiently passionate about their design to cause conflict). The subjects were provided with an on-line catalog of 60 components, grouped by type of component (e.g., volume controls), and an outline of a control panel. The MacDraw windows were arranged so the catalog occupied one half of the screen and the control panel the other half. The components could be copied from the catalog and pasted in the control panel. For the non-contentious task, subjects were provided with a list of essential components that needed to be used in the design (e.g., power control, volume control). When working in a group, the Timbuktu software was used to enable subjects to see a common version of MacDraw and to share control of the cursor. For the contentious task, subjects were assigned roles which are commonly found in industrial situations: an ergonomist, a designer, and an accountant. Each role

had 5 design principles to use in the design of the control panel. Essentially the roles encouraged the ergonomist to consider usability, the designer to consider style, and the accountant to consider cost.

## Procedure

At the start of the study the subjects were introduced to each other. An overview of the controls required to complete the design tasks in MacDraw was given. Subjects were then assigned to individual rooms.

Subjects spent a few minutes examining the communication tools. They could adjust the cameras, microphones and volume controls until they were comfortable with the quality of each other's video image and could hear each other. The operation of the shared mouse, which is available with Timbuktu software, was explained to the subjects and they experimented with it. The communication tools were then switched off until required.

Subjects were given 10 minutes to design the control panel for a stereo. This individual task was provided to allow the subject to become familiar with the design requirements for the control panel. Before proceeding with the group task the subjects read the group task instructions. In addition to designing the control panel, and to encourage discussion, the instructions told them they would need to justify their choice of component type and the location of the component on the control panel. When they were comfortable with the group task instructions, the communication medium was switched on; when communication was established, a 20 minute timer was started. Occasionally, additional elaboration of the task instructions was required for non-native English speakers.

Subjects were randomly assigned a role for the second individual task. Again, a 10 minute individual task allowed subjects to become familiar with the task and their design principles for their role. Prior to switching on the communication medium for the second 20 minute group task, subjects confirmed that they understood the goals of the task. In addition to the design principles for the role, the group task included the identification of the roles of the other two members of the group and some suggestions as to why they might not be in agreement with their principles (e.g., "The designer and ergonomist are not aware of the financial problems of the company").

Following the completion of the group task a questionnaire was completed that asked for ratings on the communication media and some personal background information.

## Measures

The measures provided a means to understand how the language and design task affected the use of the video link, and how subjects with varying language abilities compared video to audio communication.

### Video data collection

To evaluate the use of the video communication link two types of data were collected from reviewing the video tapes: the amount of time spent looking into the video monitors and the number of glances toward the monitor.

The video tapes were reviewed in order to record the number of glances toward the monitors and the total time spent looking at the monitors. These were then broken down into the number of glances and time spent looking at the monitors while talking and while simply observing.

All groups took at least 15 of the 20 available minutes to complete the group design tasks; therefore, to allow comparisons across conditions, data collected from the first 15 minutes of each design session were used.

### Questionnaire data

A questionnaire was used to assess subjects' opinions of the communication media. The questionnaire required ratings for 20 statements for both the audio and video communication conditions. (Each statement was rated on a 5 point scale.) The questionnaire covered 10 communication factors including the following problematic communication types: miscommunication, conversation repair, turn-taking, non-communication, and communication breakoff (Gass & Varonis, 1990).

## Results and Discussion

### Video results

Table II presents the group means for both the time and the number of glances. A main effect was found for the design task condition—non-contentious (NC) and contentious (C)—for all the video data sets. The subjects used the monitors significantly more in the contentious condition than in the non-contentious condition. Results between language groups showed significant differences with a greater tendency to face the monitors in the non-native speaker (NNS) and mixed (MIX) groups.

| | NS | | NNS | | MIX | |
|---|---|---|---|---|---|---|
| (Time in seconds) | NC | C | NC | C | NC | C |
| Total time spent facing video monitor | 61 | 155 | 62 | 242 | 112 | 274 |
| Time spent talking to monitor | 17 | 77 | 20 | 78 | 48 | 102 |
| Time spent observing to monitor | 45 | 77 | 42 | 163 | 65 | 172 |
| Total number of glances | 40 | 86 | 41 | 80 | 71 | 94 |
| Total number of glances to talk | 9 | 36 | 12 | 23 | 16 | 36 |
| Total number of glances to observe | 31 | 50 | 29 | 57 | 55 | 58 |

Table II Means for amount of time spent facing the video monitors, and the number of glances towards the video monitors.

Figure 2 shows a strong difference in the use of the monitors across design tasks ($f = 45.26$, $p < 0.001$) and language groups ($f = 5.25$, $p < 0.01$).

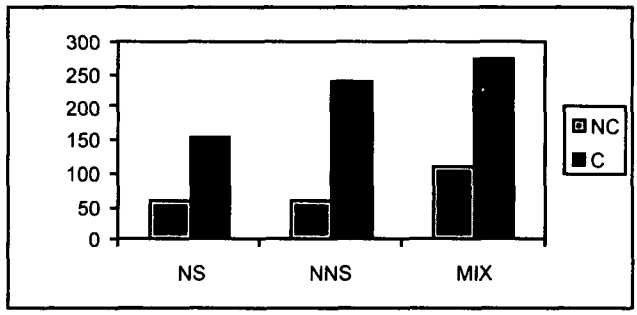


Figure 2. Total time spent facing the monitors

In considering the time spent observing the video there were main effects of design task ($f = 39.44$; $p < 0.001$) and language group ($f = 6.13$; $p < 0.01$). This was not the case for talking to the monitors. The subjects talked to the monitors more in the contentious than in the non-contentious design task ($f = 20.94$, $p < 0.001$), but there was no difference among language groups. A possible explanation could be that the NNS subjects find they need more information to interpret what they are listening to, but do not feel a greater need to speak toward other group members.

An interaction was present in the observing video data. The MIX group used the video significantly more in the contentious condition ($f = 3.94$, $p < 0.05$). This trend was also observed in the talking to video data, although it did not reach significance. The overall pattern is shown in Figure 2.

Similar effects in the time data are shown in Figure 3. The number of glances to the video monitors increase in the contentious design task condition. This was true for the summed glance data, and for the component elements of glancing to talk ($f = 22.75$, $p < 0.001$) and glancing to observe ($f = 8.20$, $p < 0.01$).

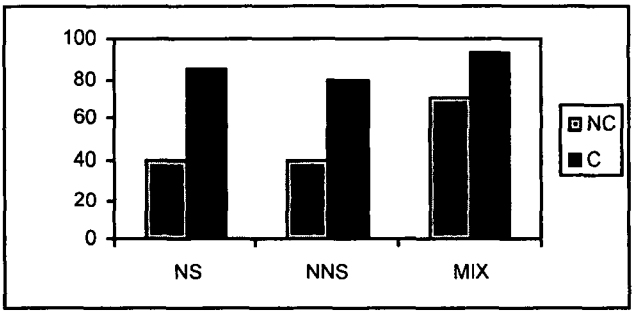


Figure 3: Total number of glances towards the monitors.

When considering the change in the use of video in accordance with the design task, the time spent facing the monitors in the contentious situation is almost 3 times that of the non-contentious condition. However, the increase in the number of glances in the contentious condition is less than twice that of the non-contentious condition. Therefore because the use of video monitors increases more than the number of glances, the average glance time increases. The duration of the glance to observe the monitors is higher in the contentious task than in the non-contentious design task for both NNS and MIX groups.

## Questionnaire results

Consistent with other studies (e.g., Olson et al., 1995; Egido, 1988), subjects rated social presence and team spirit higher in the video than the audio only condition. For three out of five problematic communication types, subjects had greater problems in the audio-only condition: miscommunication (i.e., subject felt others did not understand what subject had said), $z = -2.56$, $p < 0.01$; turn-taking (i.e., subject was interrupted), $z = -2.17$, $p < 0.05$; communication breakoff (i.e., subject gave up communicating as it was too difficult to explain ideas), $z = -2.17$, $p = 0.05$. The MIX group rated audio lower than video for all the problematic communication types: miscommunication ($z = -2.31$, $p = 0.05$); communication breakoff ($z = -1.96$, $p = 0.05$); turn-taking; conversation repair (i.e., the ease with which conversation could be repaired after a misunderstanding was noticed); and non-communication (i.e., subject not bothering too communicate as it was too difficult). (The last three showed non-significant trends).

## Discussion

The video data make it clear that video is relied upon by all language groups when communicating in the contentious situation. It is most profitable to the groups who don't consist of speakers of the same native language. Although there is some increase in the use of the video link to talk directly to others, it is of most

use in gathering additional non-verbal cues to better understand the dialog, especially in situations where conversation may be less predictable, such as when new information is presented, as in the confrontational situation.

The questionnaire data suggest that video provides positive benefits over audio in terms of the social aspect of communication and reduces the problematic communication situation, clearing the way to appreciating the content of the communication.

The MIX groups were observed to have communication problems in the design tasks with video communication, and they also self-reported that they had significant problems communicating in the audio condition. These results show the MIX group having the most difficulty in communicating regardless of the medium. Because they report having significant problems in the audio condition, it seems almost predictable that they would make the most of any enhanced channel of communication, accounting for the greater use of the video monitors. The particularly large increase in the use of the video monitors by the MIX group in the contentious task may be explained by the fact that even the NS group used the monitors significantly more in the contentious design task, whereas for the MIX group the increase in task complexity caused by the contentious task was further compounded by language communication problems.

It should be remembered that the MIX group consisted of one third native English speakers. This suggests that in mixed groups, fluency in the language being used may offer advantages, but the quality of the communication is dependent upon a joint understanding by the group. Further analyses examined differences between the behavior of the NNS subjects and the NS subjects in the MIX group. No differences were found in the non-contentious condition, but the NS subjects were found to use the video more in the contentious design task. I speculate that the difference may be related to the fact that the number of NS to NNS subjects was 1:2. The NS subject may have less experience in communicating with people speaking his/her native language as a second language and was thus experiencing communication problems not previously encountered. However, the sample size of this group is small, and further research is required to expand upon individual differences in mixed language ability groups.

A post hoc test showed a communication medium order effect for communication breakoff. The groups which received the audio condition for the non-contentious task and video for the contentious task did not report a difference in communication breakoff. However, groups that received the communication media in the order video first, then audio, report having more problems with communication breakoff in the contentious task with audio. This suggests the subjects who had received the richer communication medium for the friendly task felt a reduction in the ability to communicate ideas when the communication medium was reduced for a more complex social task, whereas the subjects who went from the poorer to the richer communication medium for the more complex

task perhaps didn't appreciate the benefits the video communication was bringing them.

This study focused on the use of the video monitors and did not measure computer screen based interactions, but a crude measure of computer use was provided by the number of components the groups had positioned in their design at the end of the experimental period. A trend suggested that with more components present on the computer screen, subjects used the video monitors less during the final 5 minutes of the experimental period. Further studies are required to explore this finding.

# Conclusion

This study addresses two aspects of video use: task contention and language ability. Although a laboratory study, it demonstrates that under certain conditions people will make significantly heavier use of a video channel. It is plausible that in workgroups showing the same characteristics, video would be more heavily relied upon. With a growing emphasis on multi-national activity, the conditions modeled in the study become more prevalent and should be investigated further.

Understanding appropriate tasks and settings for video may lead to video communication achieving some of its potential. It is hoped that this work will complement and help motivate further work on improving desktop videoconferencing quality, such as interfaces that promote a seamless transition across workspaces and maintain eye contact and awareness of gaze direction (Ishii et al., 1993; Okada et al., 1994; Ichikawa et al., 1995), and approaches that vary camera views to increase interest and awareness (Inoue et al., 1995).

Further research, including studies of longer duration and in natural work settings, is needed to confirm the findings that video is particularly helpful for groups with different priorities and linguistic backgrounds. Eventually video may find broader use; we will benefit from identifying the most promising conditions for testing and introducing the technology.

# Acknowledgments

This study was completed at Human Science Department, Loughborough University of Technology. I would like to thank Caroline Parker, Sue Joyner, and Murray Sinclair from HUSAT (Human Science in Advanced Technology) for their support and discussions on this project. Initial findings appeared in Parker, C. & Williams, G. (1994), Do you speak English? The use of telecommunications media in mixed language CSCW. *Proceedings of the 12th Triennial Congress of the International Ergonomics Association*, 1994, Vol. 4, pp. 375 - 377. I also thank Jonathan Grudin for comments and feedback on earlier drafts.

# References


Argyle, M (1969). *Social interaction* Methuen

Buxton, W. (1992) Telepresence: Integrating Shared Task and Person Spaces. Reprinted in R. M Baecker (Ed ), *Readings in Groupware and Computer-Supported Cooperative Work.* Morgan Kaufmann, 1993, pp 816–822.

Chapanis, A , Ochsman, R., Parrish, R., & Weeks, G. (1972). Studies in interactive communication· I The effects of four communication modes on the behavior of teams during cooperative problem solving. *Human Factors, 14*, pp. 487-509

Doerry, E. (1996) An empirical comparison of copresent and technologically-mediated interaction based on communicative breakdown Doctoral dissertation. Available as CIS-TR-96-01, Department of Computer and Information Science, University of Oregon, Eugene, OR.

Dourish, P., & Bly, S. (1992). Portholes· Supporting awareness in a distributed work group. *Proc CHI'92*, pp 541-547

Egido, C. (1988). Videoconferencing as a technology to support group work. A review of its failure *Proc CSCW'88*, pp 13-24

Fish, R. S , Kraut, R. E , & Chalfonte, B L. (1990). The VideoWindow System in informal communication. *Proc CSCW'90*, pp 1-11

Ichikawa, Y., Okada, K., Jeong, G., Tanaka, S , & Matsushita, Y. (1995). MAJIC videoconferencing system. Experiments, evaluation and improvement *Proc ECSCW'95*, pp. 279-292.

Inoue, T , Kobayashi, T , Okada, K , & Matsushita, Y. (1995) Learning from TV programs Application of TV presentation to a videoconferencing system. *Proc. UIST'95*, pp. 147-154

Ishii, H , Kobayashi, M , & Arita, K. (1994). Iterative design of seamless collaboration media From TeamWorkStation to ClearBoard. *Communications of the ACM, 37*, 8, pp. 83–97.

Mantei, M., Baecker, R., Sellen, A., Buxton, W., Milligan, T., and Wellman, B. (1991) Experiences in the use of a media space *Proc CHI '91*, pp. 203–208.

Magee, B. R. (1992) Enhanced factory communications (video) In CSCW'92 Video Program, SIGGRAPH Video Review Issue 87.

Okada, K , Maeda, F , Ichikawaa, Y., & Matsushita, Y. (1994) Multiparty videoconferencing in virtual social distance. MAJIC design. *Proc CSCW'94*, pp. 279-291.

Olson, M., & Bly, S. (1991). The Portland Experience: A report on a distributed research group In S. Greenburg (Ed ), *Computer supported cooperative work and groupware* London Academic Press, pp. 81–98

Olson, J S., Olson, G M., & Meader, D K (1995). What mix of audio and video is useful for doing remote real-time design work. *Proc CHI'95*, pp. 362–368.

Root, R.W (1988) Design of a multi-media vehicle for social browsing *Proc CSCW'88*, pp 25-38

Sellen, A (1992). Speech patterns in video-mediated communication. *Proc CHI'92*, pp. 49-59.

Tang, J.C , & Rua, M (1994) Montage Providing teleproximity for distributed groups *Proc CHI'94*, pp. 37-43.

Varonis, E M , & Gass, S M (1990) Miscommunication in non-native discourse. In N Coupland, H. Giles, & J. M Wiemann (Eds ), *Miscommunication and Problematic Talk* SAGE publications

Weeks, G D , & Chapanis, A (1976). Cooperative versus conflictive problem-solving in three telecommunication modes. *Perceptual and Motor Skills, 42*, pp. 879-917.

Whittaker, S (1995). Video as a technology for interpersonal communications· A new perspective. *SPIE, 2417*, pp 294-304