# Effects of Latency on Telepresence

Erik Geelhoed, Aaron Parker, Damien J. Williams, Martin Groen

HP Laboratories
HPL-2009-120

**Keyword(s):**
Telepresence, Latency, Neurological Basis of Conversation, User Studies, Empirical Methods, User and Cognitive Models, Video Analysis, Computer-Mediated Communication.

**Abstract:**
The literature on the effects of latency on mediated communication is extensive. However there have as yet not been any studies on the effects of latency in a Telepresence (Halo) setting. Round Trip latency Time (RTT) is recognized as an important determinant in helping to establish task objectives for all forms of interactive continuous media. While a number of recommendations have been made with regard to appropriate latency, these are largely based on telephone and low quality video conversations. Thus, given the provision of high quality visual information in the Halo Telepresence conferencing suites, it is necessary to explore the effect of latency on the associated user experience. In the current experiment artificially adding a delay of 250ms produced some effects, e.g. the timing of interruptions, but adding two seconds clearly affected a range of communication parameters. However, this did not produce the expected communicative breakdown, with participants persisting in their conversational style, sharing jokes, and successfully completing their task. Reviewing new developments in the research on the neurological basis of conversation as well as the role of mirror neurons, we hypothesize that the availability of (life-size) high quality video is able to buffer the potentially negative effects of latency. Thus our research effort resulted in two world firsts: [1] the evaluation of latency on Telepresence and [2] introducing mirror neuron theory into the realm of mediated communication.

External Posting Date: June 21, 2009 [Fulltext] Approved for External Publication
Internal Posting Date: June 21, 2009 [Fulltext]

# Effects of Latency on Telepresence

**Erik Geelhoed[1], Aaron Parker[1], Damien J. Williams[1] & Martin Groen[2]**

[1] Hewlett-Packard Labs, Bristol
erik_geelhoed@hp.com
[2] Department of Experimental Psychology, University of Bristol.

## ABSTRACT

The literature on the effects of latency on mediated communication is extensive. However there have as yet not been ***any*** studies on the effects of latency in a Telepresence (Halo) setting.

Round Trip latency Time (RTT) is recognized as an important determinant in helping to establish task objectives for all forms of interactive continuous media. While a number of recommendations have been made with regard to appropriate latency, these are largely based on telephone and low quality video conversations. Thus, given the provision of high quality visual information in the Halo Telepresence conferencing suites, it is necessary to explore the effect of latency on the associated user experience. In the current experiment artificially adding a delay of 250ms produced some effects, e.g. the timing of interruptions, but adding two seconds clearly affected a range of communication parameters. However, this did not produce the expected communicative breakdown, with participants persisting in their conversational style, sharing jokes, and successfully completing their task. Reviewing new developments in the research on the neurological basis of conversation as well as the role of mirror neurons, we hypothesize that the availability of (life-size) high quality video is able to buffer the potentially negative effects of latency.

Thus our research effort resulted in **two world firsts**: [1] the evaluation of latency on Telepresence and [2] introducing mirror neuron theory into the realm of mediated communication.

### Keywords

Telepresence, Latency, Neurological Basis of Conversation, User Studies, Empirical Methods, User and Cognitive Models, Video Analysis, Computer-Mediated Communication.

## 1. INTRODUCTION

Hewlett-Packard's Halo [16] Telepresence system, carefully designed in collaboration with DreamWorks Animation, provides life-size, eye-to-eye video conferencing giving participants the sense of being in the same room together. Halo features high bandwidth communication, complete soundproofing, high quality screens and cameras and, equally important, "film-set" discipline in lighting and room design, ensuring consistency and continuity. For instance, all Halo studios are identical, there is no camera movement (the screen image does not move around or change), there is no picture-in-picture, and users do not "see" the technology, they just start talking to each other [14]. Halo facilities are popular among users and preferred to traditional video teleconferencing systems. There is a general consensus that Halo meetings are lively and highly interactive with most users unable to perceive any transmission delay, however some anecdotal evidence identifies the occurrence of "simultaneous starts" typical of delayed communication. For instance, people on either side of the connection may attempt to begin to speak at (what seems) the same time, and it might take several exchanges of "No, you go ahead" before the conversation resumes its normal flow.

During the chain of signal processing and networking of all remote synchronous communication technologies a round-trip delay time (RTT) is introduced. In his keynote address to ACM Mobicom '96 Satyanarayanan stated that "…the true killer is latency" [in 5], resulting in stilted conversation full of awkward silences and accidental interruptions, and indeed participants often criticize systems for this [25]. However, there is little agreement at what point in time the latency "kills" the conversation; claims range from 100 ms to well over 600 ms [4, 15, 19, 20].

Early work on video-based communication, although lacking in estimates of RTTs, has identified certain effects of RTT. O'Conaill et al. [24] noted that interactions mediated by technologies with an inherent delay were characterized by longer turns (i.e., the time a speaker holds the 'floor' to present his or her contribution to the conversation) with an increase in formal hand-over's and reduced ability to spontaneously take the conversational floor [see also 11, 17, 35]. Although Sellen [30, 31] found no effect on measures such as pausing, overlapping speech, and interruption management, Ruhleder and Jordan [28] identified a number of additional phenomena that can arise through RTT, the causes and consequences of which, the authors claim, are not apparent to remote participants and are only evident following a micro-analysis (to 1 second accuracy) of interaction transcripts. These include words getting swapped around or comments occurring out of place, which can result in a degree of misalignment in the interaction that could lead to a potentially serious misunderstanding in the perceived meaning of a speaker's utterance.

However, no evaluations of the effect of latency on telepresence systems such as Halo have been carried out till date and these kinds of explanations may not be applicable with regard to a telepresence system. A typical telepresence and indeed face-to-face conversation does not just revolve around an exchange of intelligible utterances, but rather a temporally orchestrated, often unconscious, exchange of subtle facial expressions, eye gaze, body-language, body-stance and rhythmic sway [32], breathing [23], jaw movement, syllable and phoneme production, and oscillating brain activity [34] aimed at establishing a joint understanding of the current task objectives [3, 7].

In addition, it is well-documented that conversational partners tend to mimic one another. Proposed as the "Chameleon Effect" [6], this unconscious imitation of actions and expressions is said to improve rapport and affiliation [21], extending not just to real world scenarios but virtual ones too [1].

In this paper we report a subsection of an experiment where we artificially increased Halo's current latency and evaluated the effects on a number of self report and objective measures. Later, we will discuss Wilson and Wilsons' [34] account of a neurological basis of communication as an explanation for the occurrence of simultaneous starts in mediated communication. In addition, we relate the Chameleon Effect to the relatively recent discovery of mirror neurons and offer an explanation as to why Halo's visual quality is able to buffer adverse effects of latency.

To summarise then, this report details the **first study on the effects of latency in telepresence systems** and does so against the background of **new theoretical perspectives** in the field of mediated communication.

## 2. Method

A mixed samples, repeated measures design was employed. The within-subjects variable was latency, which had three levels: Current Halo latency, current latency+250ms and current latency+2000ms. We chose the +250ms condition because Wilson and Wilson [34] suggested that a brief silence of up to 200ms is conducive to a smooth handover during conversational turn taking. The rather extreme +2000ms condition was selected in order to severely test the participants' ability to communicate and complete the task. Although of scientific interest, we are not at liberty to disclose Halo's current latency. The three latency conditions were presented in one of two orders, the between-subjects variable: Order 1, current latency to longest latency and Order 2, longest latency to current latency. Latency was artificially increased without affecting lip-synchronization. The dependent variable was participants' perceptions of the Halo experience, which was assessed through a questionnaire and a discussion at the end of the session. A frame-by-frame analysis of a random sample of the video footage provided additional data.

### Participants, Task, Procedure

Using two Halo studios 32 participants (11 females, 21 males, age range 25 - 60) took part in the study: 16 participants in HP Bristol, U.K. and 16 in HP Corvallis, U.S.A. A total of eight experimental sessions involved two pairs of participants, one pair in each location, and lasted approximately 50 minutes.

Three similar tasks were employed which required participants to work co-operatively in order to generate and prioritize items describing either an ideal car, home, or holiday/vacation. The topics were chosen because they are subjects that most people can (and like to) talk about without preparation, quickly generating a good common ground between the participants. Each of the tasks was paired with a specific latency condition/setting:

1. Ideal car – current latency
2. Ideal home – current latency+250ms
3. Ideal holiday/vacation – current latency+2000ms

Each latency condition lasted for 10 minutes, with a similar procedure across the three conditions. Participants were instructed that during the first five minutes, they should work together to generate at least seven items that described an ideal home, car, or holiday/vacation. For the remaining five minutes the participants were required to work together to prioritize the top seven of the items they had generated. The experimenter was on-hand to inform the participants when the five minutes were up. After the 10 minutes were up participants completed a continuous, graphic rating scale questionnaire [33] containing 11 items, enquiring about their perceptions of verbal and non-verbal aspects of the conversational flow:

1. How productive did you find this Halo session?
2. To what extent did you notice a delay in the communication, i.e. it looked like the people in the other Halo room heard you a little while after you spoke?
3. How disruptive did you find the delay?
4. How easy was it to keep track of the dialogue?
5. How often did it happen that someone in your Halo room and someone in the other Halo room started talking at the same time?
6. How lively were your discussions in the Halo room?

7. How easy was it to take the floor in the Halo session?
8. How easy was it for the others to take the floor in the Halo session?
9. How natural did it feel when you interrupted someone talking in the other Halo room?
10. How natural did it feel when you were interrupted when talking?
11. How often were there awkward silences between the two Halo rooms?

At the end of the experimental session participants were given the opportunity to make comments about the experience. Sessions were recorded using the Halo cameras and microphones.

## 3. Results

The questionnaire was analyzed using an Analysis of Variance (ANOVA), where the more conservative Type IV sum of squares correction was applied due to unequal numbers for the factor "Order". An alpha level of .05 was adopted. To save space and improve clarity for each of the outcomes only p-values and Cohen's [10] partial eta-squared ($\eta_p^2$) are provided. More detailed results can be requested from the authors. In addition, similarities between the questions were investigated using Multi-Dimensional Scaling. These latter analyses helped us to understand which aspects of communication were affected by an increase in latency.

### Questionnaire

Table 1 lists the results of the 2-Way ANOVA (Latency * Order) of the questionnaire data, which were all highly significant, ordered by the size of $\eta_p^2$. Thus, across the three conditions the largest effect was due to "being interrupted" and the smallest effect was a result of how lively the session had been. The liveliness of the session was also affected by the factor order, p = .016 (see Figure 4).

| Question | Latency | | Order | | Current Vs +250ms |
|---|---|---|---|---|---|
| | p | $\eta_p^2$ | p | $\eta_p^2$ | |
| Being interrupted | <.001 | .623 | | | ** |
| Notice delay | <.001 | .523 | | | ** |
| Interrupt | <.001 | .504 | | | * |
| Disruptive | <.001 | .488 | | | * |
| Keep track | <.001 | .476 | | | |
| Take floor | <.001 | .468 | | | |
| Others take floor | <.001 | .408 | | | ** |
| Awkward silence | <.001 | .331 | | | |
| Productive | <.001 | .264 | | | |
| Simultaneous start | <.001 | .236 | | | |
| Lively | .003 | .177 | .016 | .179 | |

**Table 1: Summary main effects in order of $\eta_p^2$**

Tukey's post-hoc analysis revealed that the difference between the current and the current+250ms condition for five questions can be attributed to the latency manipulations. This is signified by the stars in the last column of table 1 (* = p < .05,

** = p <.01). Tukey’s post hoc comparisons between the +250ms and the +2000ms (not shown here) were all significant. It is apparent that for all these self-report measures adding two seconds of RTT has an undeniable *and* differential effect.

As examples we show the results for “Noticing the delay” and “How natural it feels being interrupted” (both displayed in figure 1).

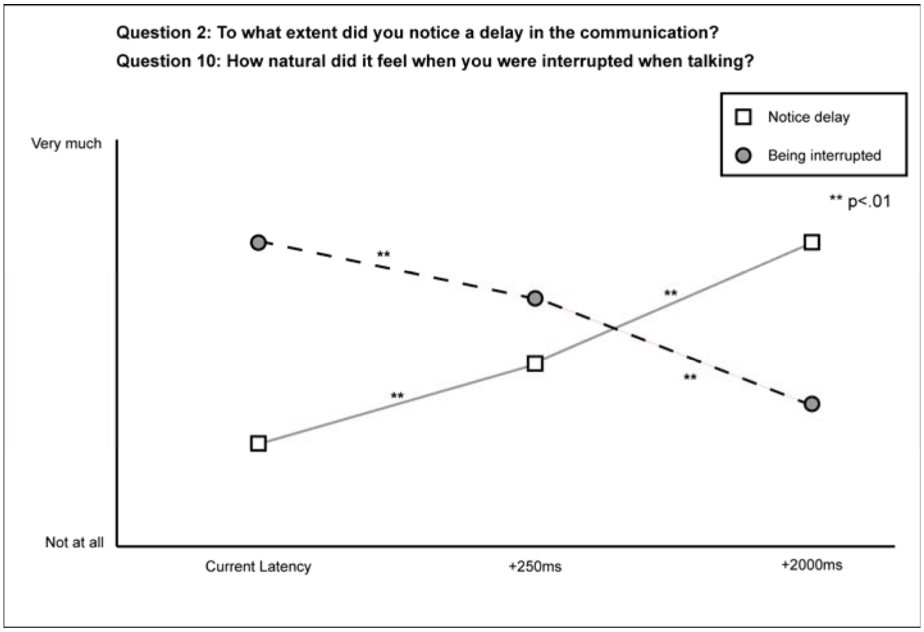


**Figure 1: Example results Noticing delay and Being interrupted.**

Thus, in the current latency condition participants did not appear to notice the delay, but adding 2000ms resulted in the delay being very noticeable. Similarly, in the current latency condition, being interrupted was reported to feel quite natural, and a lot less natural in the +2000ms condition. For both questions the +250ms maintains an intermediate position. Note that as a result of the difference in phrasing of the two questions, the first one (noticing the delay) results in the rising of the mean score as latency increases whereas for the second (how natural it feels being interrupted) the means decrease as latency increases. In the questionnaire there were seven questions phrased in such a way that scores decreased with latency increasing and four that increased. This was done to prevent the participants to respond in an ‘automatic pilot’ mode.

This is also reflected in the Multi Dimensional Scaling (MDS) analyses (see Figure 2 for the solution for the +2000ms condition). In the two left quadrants are those four items resulting in increasing means (signifying poor communicative flow) and on the right-hand side those where the means decreased (signifying good flow). Furthermore, the items in the bottom two quadrants were most affected by high latency while those in the top right quadrant were least affected. The numbering of the items refers to the order of the means (with 1 indicating the highest mean).

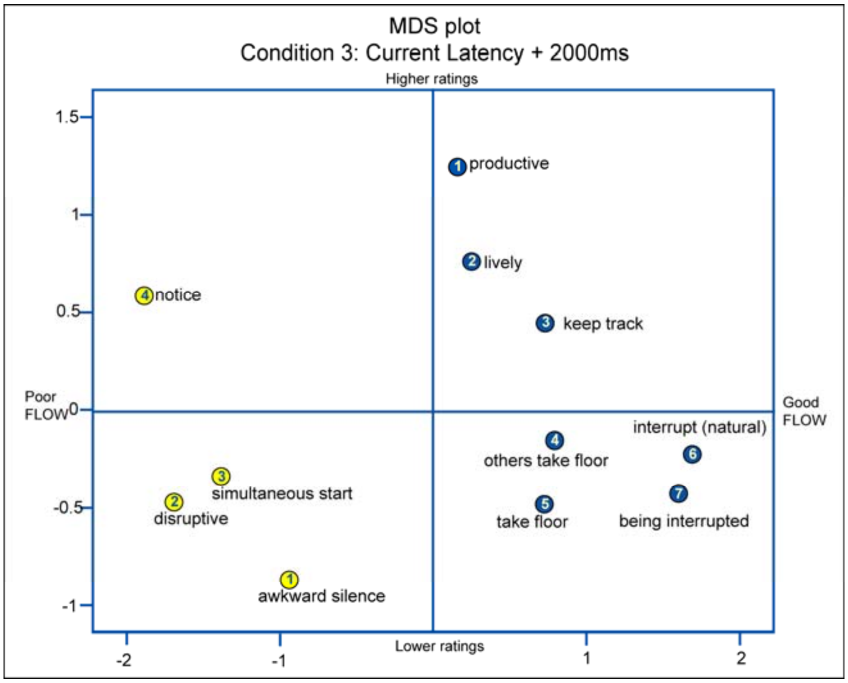


**Figure 2: MDS plot for +2000ms condition**

Figure 3 shows the effects of latency on the responses of the whole questionnaire for the three RTT conditions with the means for the four negatively phrased questions reversed. It is clear that in the current latency condition most means are high (around 75%), with the exception of simultaneous starts. The results from the +2000ms condition showed that certain items are less affected (that is, how productive and how lively the sessions were) than others, in addition to noticing the delay, interruptions seem to suffer and awkward silences as well as simultaneous starts begin to hamper the communication.

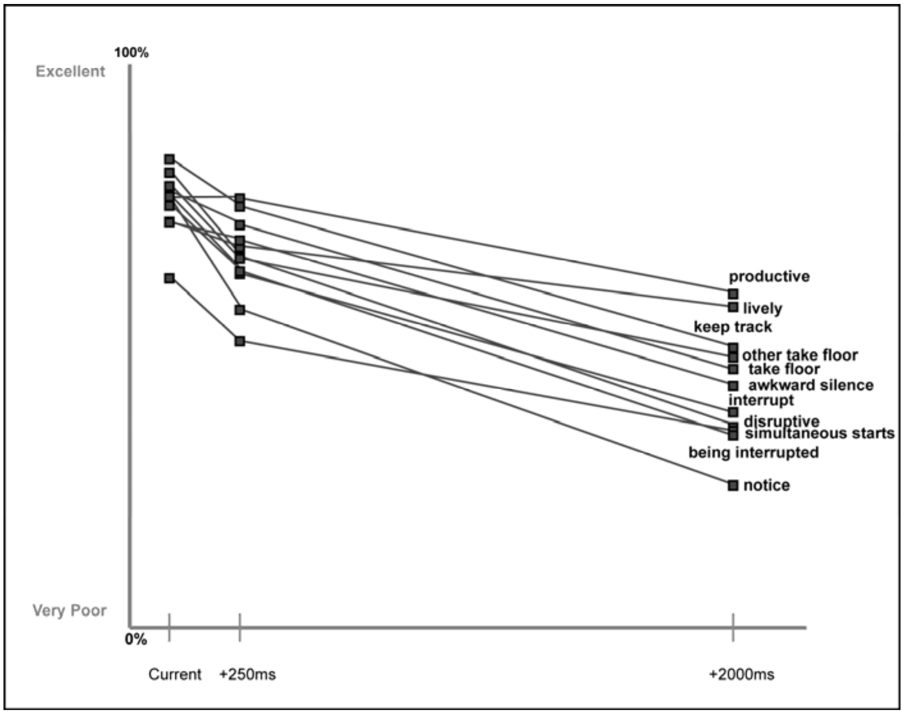


**Figure 3: Effects of Latency across conditions**

The only effect of the order of the presentation of the latency conditions was on how lively participants reported the sessions to be. Figure 4 shows that when participants started with the current (shortest) Halo latency and ended with the longest (Order 1) the liveliness of the discussions was rated significantly less ($p < .05$) than for those who started with the longest latency and ended with the shortest (Order 2)

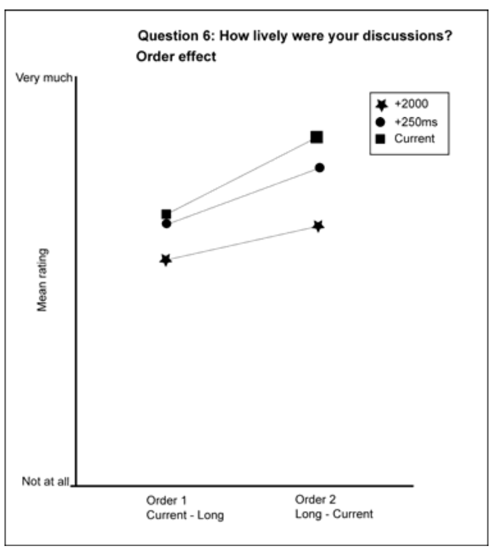


**Figure 4: Order effects for Liveliness**

## Video analysis

A frame-by-frame (1 frame = 40ms) audio-video recording analysis was conducted on the speech of the current and +2000ms conditions of four sessions: Sessions 3 and 5 in the Order 2 condition and 6 and 10 in Order 1. Although less precise than audio-recording analysis, we were better able to identify lip movement as an indication of when someone started or ended talking or laughing. Table 2 lists some of the results.

| Session & Order | | 3 Order 2 | 5 Order 2 | 6 Order 1 | 10 Order 1 |
|---|---|---|---|---|---|
| Silence | current | 26% | 14% | 32% | 20% |
| | +2000ms | 33% | 17% | 30% | 40% |
| Speech overlap between sites | current | 48 | 68 | 19 | 26 |
| | +2000ms | 59 | 89 | 30 | 19 |
| Simultaneous starts between sites | current | 22 | 24 | 4 | 8 |
| | +2000ms | 15 | 24 | 7 | 7 |
| Laughter between sites | current | 6 (6) | 4 (5.5) | 11 (9.75) | 3 (6.75) |
| | +2000ms | 0 (3.5) | 4 (6) | 5 (5.75) | 7 (8.75) |

**Table 2: Some AV analysis results**

The percentage of "silence" varied from session to session and although for one session (10) there was a 100% increase in the amount of silence between the current and +2000ms conditions, the other three sessions did not follow this pattern. Thus, the silence is not just a result of waiting for each other to finish or start talking. With an increase in latency, three sessions (3, 5 and 6) showed an increase in the frequency of speech overlap between the two sites but the fourth showed a decrease. Simultaneous starts were tentatively defined as at least one of the dialogue partners starting to talk within 10 frames (=400ms) of each other and again no discernable pattern was evident. There were six bouts of joint laughter for session 3 between the two sites in the current latency condition out of a total of 24 laughter-turns (i.e. between the four participants a mean of 6, shown in parentheses in the table) and none in the +2000ms condition (average laughter-turns 3.5). On the other hand there was more joint laughter for session 10 in the +2000ms condition, even though there was more silence for this session.

In other words the results are too variable to be able to draw conclusions about the effects of RTT increase on speech patterns. An investigation of individual patterns indicated that participants were consistent in the way they dominated or did not dominate the sessions but there was no evidence of a consistent increase in turn length as RTT increased. Coupled with the fact that participants completed their tasks within the time limit, we found no consistent differences between the effects of latency on participants' conversational styles, sharing jokes and task strategy.

## Comments afterwards

From the discussion afterwards it appeared that participants did not clearly identify the +250ms condition as different from the current RTT one:

*"I think the second was just about noticeable but didn't interfere too much, but the third one did interfere"*
*"if you hadn't told me I probably wouldn't have known"*

The conversation in the +2000ms condition was identified by participants as considerably inhibiting the interaction:

*"I think it slowed the conversation down, I think our efficiency went"*

Some of the comments made by participants indicated that they were able to work around the effect of latency by modifying their behaviour:

*"there were certain times when I did modify what I was doing because I knew there was a delay yeah"*
*"I think we adapted"*

## 4. Discussion

The current experiment investigated the effect of adding 250ms and 2000ms to Halo's current latency on interaction between two distributed sites. In the current latency condition the lowest rating was for "simultaneous starts" but by and large the other questions were rated high. Although the self-report measures clearly showed negative effects of adding 2000ms, there were fewer significant results for the increase of +250ms (mainly noticing the delay and its effects on interruptions) than expected: It would appear that the design of the Halo studio and the visual quality of the cameras and screens is able to buffer the adverse effects of modestly increased latency. In addition, it is puzzling that with the added +2000ms RTT communication did not break down, people completed their task, with conversational embellishments such as joke telling remaining intact. This was also reflected in the relatively higher scores for sessions being productive and lively in the +2000ms condition.

MDS plots are based on correlations and provide a good way to show similarities and differences in one graphic [36]. Inspecting the Y-axis and the right hand side of the plot, we see at the top with the highest mean rating the item "productive", further down there is the second highest mean for "lively", then below that in the third place "keep track" and so on, although the items in the 6th and 7th place follow a slightly different pattern. On the left hand side starting at the bottom and then going up we see a similar pattern, also in order of mean rating. However it is very unusual for MDS to reflect the order of the mean ratings. Or rather: Having used MDS as a tool to show similarities for twenty years, we have never come across this phenomenon before. We speculate that this somehow reflects the orchestrated, or at least ordered, nature of communication (e.g., as in Clark's [9] action ladder of completion).

Participants starting in the longest latency condition rated the liveliness of the sessions increasingly higher as their latency condition got shorter, possibly indicating some sort of relief, whereas those who started with the shortest latency showed only a slight drop in their ratings as RTT increased. It is also noteworthy that even beginning with that +2000ms condition, participants' ratings were still above average, confirming that even under the extreme condition of +2000ms sessions were still much livelier than one would expect.

How can these findings be explained?

In the following section, we review the relevant literature on mimicry and mirror neurons and link these to our findings.

In addition to the aforementioned Chartrand & Bargh's [6] "Chameleon Effect" where conversational partners have a tendency to mimic one another, we can see a potential underlying imitation system, that of the mirror neuron. This system, present within the pre-motor and inferior parietal cortices, contains neurons that activate not only when an action is performed but also when it is perceived. In fact, it is not just overt actions that causes activation in this system, facial emotions and hand actions [22] as well as speech [12] have also been shown to have overlapping areas of activation for action and perception and research in other domains are reporting similar findings. This in turn has provided evidence in favor of motor-based theories of comprehension in a number of domains, particularly speech where it has been hypothesized that the mirror neuron system may represent the network from which speech evolved [26]. Although mirror neuron activation does not normally lead to actualization of imitation, there appears to be some role of the spinal cord in preventing this [2], it is claimed that it may prime mimicry to some extent [27].

Drawing us to tie these two together is the role of empathy that has found to be both a facilitator of the Chameleon Effect, and, a predictor of the strength of mirror neuron activation in response to facial expressions [18] as well as sounds [13]. In presence research too, it has been shown that those with higher empathy ratings, as shown by self-report questionnaires, experience greater sensations of co-presence [29]. If we assume that co-presence is associated with empathy, which in turn is a product of mirror neuron activation and mimicry, then it seems that these media that support a richer degree of sensory information (telepresence) help to create not only a sense of co-presence but also facilitate a degree of grounding [8] and comprehension. Thus we argue that our findings can be interpreted as an indication that telepresence systems offer better facilitation to those mechanisms relating to the mirror neuron system and mimicry in comparison with lower quality systems.

In addition, we found no effect of latency on turn-taking in Halo sessions. Wilson and Wilson's account of a neural conversational turn taking mechanism [34] could provide an explanation.

In face-to-face communication, when someone is speaking, a listener will anticipate with minute precision when a conversational turn is ending, implying an entrainment of timing between participants in the conversation. This suggests some form of cyclic patterning, an involvement of endogenous oscillators in the human brain, which are populations of neurons that collectively show periodicity serving timing-related activity. Brain-based oscillators show properties of what are called relaxation oscillators (in contrast to harmonic oscillators) and are susceptible to outside influence only during one phase of their cycle and therefore, are highly stable, robust, and predictable in their timing properties.

Wilson & Wilson explain that there is a close link between this fine-grained high frequency oscillation and the rhythmic production of phonemes. However, speech-rate, syllable production, which occurs at a larger time scale, seems less rhythmic. This is explained by another known phenomenon where lower frequency (more irregular) oscillators fine tune stable higher frequency oscillation. In this way, features of communication, whether fine- or course-grained, all play their role and work in concert.

In addition, Shockley et al. [32] point out the importance of, and effect on, cooperative conversation of body posture and sway. Nodding head movements and (hand and arm) gestures are well recognized as being an integral part of (lively) conversations. Breathing is another one of those lower frequency and coarser oscillators that helps with smooth turn taking in conversation. Failure to coordinate breathing around a turn transition is associated with simultaneous starts of speech [23].

Wilson and Wilson propose that speaker and listener are engaged in a counter-phase locked neuronal activity, whereby both participants in the conversation anticipate the end of a turn, accompanied by a brief silence of about 200ms, which is conducive to a smooth handover as well as avoiding a simultaneous start. If the silence is longer, e.g. through RTT, then phases unlock somewhat and mutuality is attenuated. These longer silences make the likelihood of a simultaneous start higher as both parties in a conversation as both have the opportunity to start.

Thus, we argue that the overall design of the Halo studios (identical rooms, soundproofing, good quality audio, lighting and superior image quality) is conducive to perceiving facial and body "language". We suggest that this permits the activation patterns of participants' conversational neural oscillators to become entrained as well as supporting the Chameleon effect (through appropriately firing mirror neurons). This enhances the perceived co-presence of remotely located participants, and may help negate the effects of the delay. It appears that providing a life-like conversational experience, such as with Halo, could mitigate some of the problems that were observed with mediated communication of a lower quality.

## ACKNOWLEDGMENTS

We thank Stephen Albright, Tim Hubley, Mike Derocher, Mark Gorzynski, and Brad Allen of the HP Halo R&D in Corvallis, and Jan Noyes of the University of Bristol. Aaron Parker is supported by Great Western Research, who are gratefully acknowledged.

## REFERENCES

1. Bailenson, J.N. and Yee, N. Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science, 16,* 10 (2005), 814-819.
2. Baldissera, F., Cavallari, P., Craighero, L. and Fadiga, L. Modulation of spinal excitability during observation of hand actions in humans. *European Journal of Neuroscience, 13*, 1(2001), 190-194.
3. Bangerter, A. and Clark, H.H. Navigating joint projects with dialogue. *Cognitive Science, 27*, 2(2003), 195-225.
4. Brady, P.T. Effects of transmission delay on conversational behavior on echo-free telephone circuits. *The Bell System Technical Journal, January,* (1970), 115-134.
5. Cheshire, S. *Latency and the quest for interactivity*. White paper commissioned by Volpe Welty Asset Management L.L.C for the Synchronous Person–to-person Interactive computing Environments Meeting. San Francisco (1996).
6. Chartrand, T.L. and Bargh, J.A. The Chameleon Effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 6 (1999), 893-910.
7. Clark, H.H. and Wilkes-Gibbs, D. Referring as a collaborative process. *Cognition*, *22*, (1986), 1-39.
8. Clark, H.H. and Brennan, S Grounding in communication. *Resnick, L.B. Levine, J. and Teasley, S.D. Perspectives on Socially Shared Cognition,* APA Press, Washington, (1991).
9. Clark, H. H., Using language. New York: Cambridge University Press, (1996).
10. Cohen, J. Eta-squared and partial eta-squared in fixed factor anova designs. *Educational and Psychological Measurement, 33*, 1(1973), 107-112
11. Cohen, K. (1982). Speaker interaction: Video teleconferences versus face-to-face meetings. In *Proc. of Teleconferencing and Electronic Communications,* University of Wisconsin Press, (1982), 189-199.
12. Fadiga, L., Craighero, L., Buccino, G. and Rizzolatti, G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience , 15,* (2002), 399-402.

13. Gazzola, V., Aziz-Zadeh, L. and Keysers, C. Empathy and the Somatotopic Auditory Mirror System in Humans. *Current Biology,16,* 18 (2006), 1824-1829.

14. Geelhoed, E., Williams, D.J., Albright, S. and Hubley, T. User-Centred Approach to Achieving the Halo Experience, *eChallenges e-2007 Conference,* (2007).

15. Hammer, F., Reichl, P. and Raake, A. The well-tempered conversation: Interactivity, delay and perceptual VoIP quality. *Proc. of the 2005 IEEE International Conference on Communications* (2005), Vol. 1, 244- 249.

16. HP Halo Telepresence and video conferencing solutions http://www.hp.com/halo/

17. Isaacs, E. and Tang, J. What video can and can't do for collaborations: A case study. *Proc. of the ACM Multimedia '93 Conferen*ce, ACM Press, (1993).

18. Jabbi, M., Swart, M. and Keysers, C. Empathy for positive and negative emotions in the gustatory cortex. *NeuroImage, 34,* (2007), 1744-1753.

19. Klemmer, E. Subjective evaluation of transmission delay in telephone conversations. *Bell System Technical Journal, 46*, (1967), 1141-1147.

20. Krauss, R.M. and Bricker, P.D. Effects of transmission delay and access delay on the efficiency of verbal communication. *The Journal of the Acoustical Society of America, 41*, 2 (1967), 286-292.

21. Lakin, L.J. and Chartrand, T.L. Using non-conscious behavioral mimicry to create affiliation and rapport. *Psychological Science, 14,* 4 (2003), 334-339.

22. Leslie, K.R., Johnson-Frey, S.H. and Grafton, S.T. Functional imaging of face and hand imitation: towards a motor theory of empathy. *NeuroImage, 21*, (2004) 601-607.

23. McFarland, D.H. Respiratory markers of conversational interaction. *Journal of Speech, Language & Hearing Research, 44,* (2001), 128-143.

24. O'Conaill, B., Whittaker, S. and Wilbur, S. Conversations over video conferences: An evaluation of the spoken aspects of video-mediated communication. *Human-Computer Interaction, 8,* (1993), 389–428.

25. Olson, G., & Olson, J. (2003). Mitigating the effects of distance on collaborative intellectual work. Economics of Innovation and New Technology, 12(1), 27-42.

26. Rizzolatti, G. and Arbib, M.A. Language within our grasp. *Trends in Neurosciences, 21,* 5 (1998), 188-194.

27. Rizzolatti, G. and Craighero, L. The mirror-neuron system. *Annual Review of Neuroscience, 27*, (2004), 169-192.

28. Ruhleder, K. and Jordan, B. (1999). Meaning-making across remote sites: How delays in transmission affect interaction. In S. Bødker, M. Kyng, and K. Schmidt (Eds.), *Proc. of the Sixth European Conference on Computer Supported Cooperative Work: ECSCW 99*. 12-16 September (1999) Copenhagen, Denmark.

29. Sas, C. and O'Hare, G.M. Presence equation: An investigation into cognitive factors underlying presence. *Presence, 12*, 5 (2003), 523-537.

30. Sellen, A.J. Speech patterns in video-mediated communication. In *Companion Proc. of CHI'92 Human Factors in Computing Systems,* ACM Press, (1992), 49-59.

31. Sellen, A.J. Remote conversations: The effects of mediating talk with technology. *Human-Computer Interaction, 10*, (1995), 401–444.

32. Shockley, K., Santana, M.V. and Fowler, C.A. Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 2(2003), 326-332.

33. Stone, H., Sidel, J., Oliver, S., Woolsey, A. and Singleton, R.C. (1974). Sensory evaluation by quantitative descriptive analysis. *Food Technology*, Nov (1974), 24-34.

34. Wilson, M. and Wilson, T.P. An oscillator model of the timing of turn-taking. *Psychonomic Bulletin and Review, 12*, 6 (2005), 957-968.

35. Whittaker, S. Rethinking video as a technology for interpersonal communication: Theory and design implications. *International Journal of Human-Computer Studies, 42*, 5 (1995), 501-529.

36. Young, F.W. and R.M. Hamer, *Multidimensional Scaling: History, Theory, and Applications*. Lawrence, Erlbaum Associates, London, (1987).