How Text and Audio Chat Change the Online Video Experience

Justin D. Weisz
Computer Science Department
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh, PA 15213
jweisz@cs.cmu.edu

Sara Kiesler
Human Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Ave.
Pittsburgh. PA 15213
kiesler@cs.cmu.edu

ABSTRACT
Many online video sites provide a text chat feature so viewers can chat with others while watching videos. How does chatting affect their experience? Would audio chat be more fun or would it be too distracting? The richer medium of audio may more closely approximate the living room or club experience, but human factors research suggests that audio chat could increase distraction and potentially detract from the viewing experience. This paper presents the results of an experiment comparing text with audio chat when the video does or does not have dialogue, and when viewers are watching the videos in the same or a different order. A control group watched videos without chat. Overall, audio chat and text chat were equally distracting, and chat was more distracting when the video contained dialogue. Despite the presence of distraction, viewers found both text and audio chat enjoyable. Those who used audio chat preferred it to text chat for talking with others while watching videos with their friends.

Categories and Subject Descriptors
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization interfaces – *synchronous interaction*

General Terms
Design, Experimentation, Human Factors

Keywords
Video, interactive TV, social TV, audio chat, text chat, friends, synchronized video, streaming, playlist

1. INTRODUCTION
There exists a dichotomy in the social norms surrounding interaction while consuming media. In our living rooms, and in bars and clubs, we can enjoy watching television shows and movies in the noisy company of friends, family or strangers. In these social settings, fellow viewers tolerate or encourage comments and conversation with others. The sense of companionship may have priority over the viewing experience. In less interactive settings, such as in movie theatres, viewers want to be engaged in the video, and do not like to hear others talk. Movie theatres often play a video reminding the audience to turn off cell phones and not to speak to neighbors during the show.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference'04, Month 1–2, 2004, City, State, Country.
Copyright 2004 ACM 1-58113-000-0/00/0004…$5.00.

Online video, dominated by short amateur clips, has tended to follow the living room model, encouraging social interaction with video watching. Most online video sites experimenting with real-time interaction have added a text chat feature for their viewers. For example, UStream.TV provides text chat side-by-side with streaming video (see Figure 1).

Figure 1. Example of video with text chat.

Alternatively, viewers can open up IM or chat on their computers, independent of the online video system. For example, Zync is a plugin for Yahoo Messenger that lets people coordinate watching YouTube videos in an IM session [18]. Further, a few television-based systems have been created that allow viewers to interact with one another using a text chat feature [1][2][13][16]. Others allow viewers to chat with each other using predefined "canned" messages [8].

Recent research in interactive and social TV suggests that text chatting while watching online videos is enjoyable. When chat is optional (i.e., when viewers can close, hide or ignore the chat window), viewers uninterested in chatting should not feel imposed upon by the chat. For those users interested in chatting, research in human factors suggests that chat will be distracting. Reading and typing draw the viewer's visual focus away from the video, and while cognitively processing chat conversation, the viewer has fewer mental resources to cognitively process the video content [22]. Despite this distraction, research thus far suggests that the degree of distraction present when using a text chat feature while watching a video is comparatively small, and not sufficient to reduce the enjoyment of the video [21]. Admittedly,

researchers testing distraction so far have not shown participants complicated or deep videos that require concentration.

1.1 Audio Chat with Online Video?

Audio chat is another option for online video, but it is still comparatively rare. One system that uses audio chat with TV is AmigoTV [3]. AmigoTV is an interactive television system that combines voice chat with an overlay display on the TV display showing avatars of other viewers.

There are several reasons to offer viewers an audio chat option. Talking and listening, on average, is a richer and more engaging experience than is reading and writing [4][5]. Having an audio chat with others could make the online video experience more like the living room or club experience, as well as make chat accessible to people who are uncomfortable with technology or are unable to use computer keyboards. Scholl et al. [17] compared text and audio chat in a multimedia conferencing system, and they found that audio chat was easier to use and felt more natural than text chat for communicating feelings and emotions.

Audio chat is likely to be more distracting than text chat. If audio chat is especially engaging, it will draw attention away from the video and cause viewers to lose track of the content. Another reason audio chat is likely to be more distracting than text chat is that people do not easily process two speech signals at once when both are heard simultaneously [11][14][15]. The audio signal stays in mind for only a few seconds, so when there are two competing audio signals, listeners have to switch back and forth to process both. This switching effort will compete with cognitive processing of watching a video. Thus, audio chat is likely to interfere with video dialogue (if dialogue is present), reducing understanding and memory of the video content.

This paper reports a study in which we compared audio chat with text chat and no chat under different video content conditions. The purpose of the study was to examine the comparative advantages and disadvantages of the two kinds of chat with different video content, so that designers of systems can begin to assess how chat features are likely to affect the viewing experience.

1.2 Streaming vs. Playlist Model

Online video sites that provide a chat feature to their members follow one of two models for how viewers watch videos. In the case of sites such as Justin.TV, Lycos Cinema and UStream.TV, viewers watch the same video at the same time. This is a *streaming model*, as viewers watch a live (or simulated "live") video stream together. In this case, the content seen by viewers is synchronized (although network delays and jitter may cause small asynchronies in playback over time). Figure 1 depicts a typical interface for a site that uses the streaming model – here, viewers chat with each other while watching a web comic artist draw his latest comic.

In the *playlist model*, viewers pick and choose which content they want to watch, independent of other viewers. When chat is integrated into the playlist model, viewers join a chat room or a channel to watch videos. Each channel typically has its own theme or topic as well as its own library of video content. Once viewers have entered a channel, they can browse the list of available videos and build a personal playlist of videos they want to watch. These videos may be different from what other viewers in the channel are watching. However, all viewers in the channel can chat with each other in real time. Sites such as YouTube Streams and Gaia Online, and systems such as Joost use a playlist model.

In this paper, we also examine the differences between the streaming and playlist models of video playback in terms of their ability to sustain audio and text chat.

1.3 Research Questions

1.3.1 Do viewers enjoy using audio chat?

Audio chat may be more enjoyable than text chat because of its ease of use and its ability to convey richer emotional content. On the other hand, audio chat may be more distracting than text chat if viewers are watching a video as well as talking. We predict that, regardless of its impact on distraction and enjoyment of the videos, audio chat would be more a enjoyable form of socializing than text chat because audio is a richer medium for social interaction.

1.3.2 Does distraction from chat reduce enjoyment?

Because chat with video inherently requires multitasking, we need to understand whether any resulting distraction affects people's enjoyment of the online video experience. As noted above, the human factors and attention literature suggest that both text and audio chat will be distracting. However, this literature is primarily focused on task performance and does not address people's enjoyment of the video experience. Many people, especially young people, enjoy combining media use with another non-media activity [10]. Media use might reduce social anxiety and self-consciousness when talking with others because the video takes peoples' attention away from themselves.

Our previous study of text chat while watching online video [21] suggests that the distraction from text chat and enjoyment are independent experiences – chat does not detract from the video experience. However, that study only tested the effect of a text chat feature, and only on the enjoyment of videos without dialogue. In this study, we examine the association of distraction and enjoyment when viewers talk with others through audio chat or text chat, and we examine the impact of video content with and without dialogue.

The human factors and attention literature [22][11][14][15] suggests that audio chat will be particularly distracting when the audio channel is overloaded (audio chat with heavy auditory content). We predict that audio chat will be more distracting when viewers are watching videos with dialogue.

1.3.3 Can text and audio chat be sustained without a shared media context?

Our final question compares the streaming model with the playlist model. In the streaming model, viewers are synchronized with respect to the content they watch, and thus they share a media context with each other. A viewer who laughs and says "that gopher is so funny" can be secure in the knowledge that his fellow viewers will understand who "that gopher" is.

In the playlist model, unless viewers make the effort to watch the same video, they are unsynchronized and lack a shared media context. Because they are seeing different videos, it will be more difficult for them to chat with each other about what they are seeing. They will need to put more effort into determining suitable topics for chat. We predict these viewers will chat less about what they are seeing, and more about personal topics or other sources of commonality (e.g., their job, school, or location).

We expect audio chat to be particularly difficult to use in a playlist model. Because viewers lack a shared media context, they cannot coordinate their speaking to quiet times in the video. Therefore, audio chat should be particularly distracting in a playlist setting for videos containing verbal dialogue.

2. TEXT VS AUDIO STUDY

We ran a controlled laboratory study to answer our research questions about how text versus audio chat leads to distraction and enjoyment, attitudes towards audio chat, and the differences between the streaming and playlist models.

2.1 Method

Groups were recruited for the study using an experimental scheduling website. They came to a laboratory room to watch a series of short videos on the computer. Each group consisted of three people who knew each other before the study and considered each other as friends.

2.1.1 Experimental design

The experimental design was a 4 x 2 x 2 factorial design comparing four chat conditions (no chat vs. text chat vs. audio chat vs. both text and audio chat), video synchrony (videos in the same order vs. different order) and dialogue presence (no dialogue vs. dialogue). Chat conditions and video synchrony were between-subjects factors and dialogue presence was a within-subjects factor.

2.1.2 Participants

Participants were recruited in groups of three friends from the psychology experiment directory at Carnegie Mellon and on-campus fliers. Forty-eight groups were recruited, for a total of 144 participants. Group assignments to each condition were random and equally balanced – 6 groups were assigned to each chat and video synchrony combination.

The average age of the participants was 23.8 years (SD = 7.2 years); 52 participants (36%) were female. Eighty percent of participants were students (44% graduate, 36% undergraduate). Twenty percent reported other affiliations such as alumni or visiting scholar, or did not list their affiliation. Participants were paid $15 each for their participation, which took approximately one hour.

2.1.3 Procedure

Participants were informed that they would watch a series of videos on the computer and take a survey at the end. Participants were seated in separate rooms and could only communicate with each other using the methods we provided.

Groups with text chat could type messages to one another using web-based chat software (see Figure 2). Groups with audio chat could speak to one another using headsets. For audio chat, we ran the TeamSpeak software (http://www.goteamspeak.com/) in the background and used the voice-activation feature so participants could speak to each other without having to press a key on the keyboard. Audio levels were tested and adjusted before the beginning of the experiment to make sure everyone could hear each other properly. Participants with both text and audio chat were told that they could use either method for communicating with their friends.

The videos used in the study were chosen from highly rated YouTube videos that were between three and seven minutes long. Two types of YouTube videos were selected for use. To create the

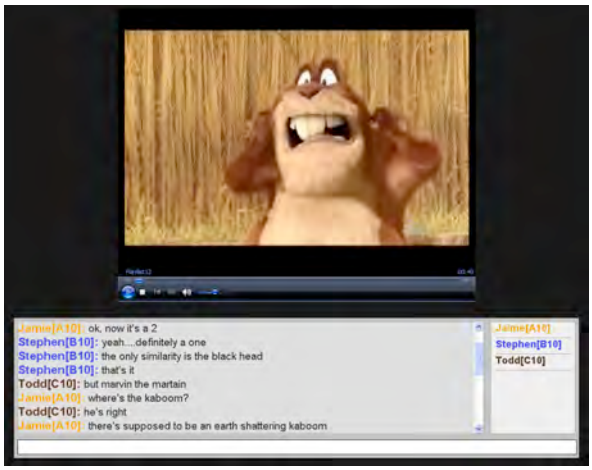
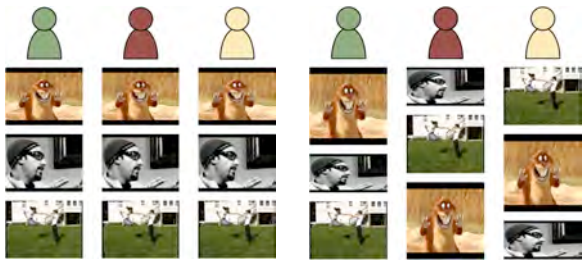


Figure 2. Screenshot of the text chat condition.

no dialogue condition, four videos contained no auditory dialogue, but did contain musical soundtracks providing an auditory backdrop for the videos. To create the dialogue condition, four videos included talking either by the characters or a narrator. These two types of videos allowed us to compare the effect of chat on different types of video content.

We used Windows Media Encoder to stream the videos to each participant's computer. To manipulate video synchrony, groups in the same video order condition watched the eight videos in sync with each other, such that they saw the same video content at the same time. Members of groups in the different order condition each watched the videos in a different, random order (we used 3 media encoder instances to achieve this). Figure 3 depicts the difference between the same order and different order groups. Note that the transition points from one video to another do not necessarily coincide for participants watching in a different order. Because it was possible that a randomized ordering would result in some participants having the same initial video, and therefore begin with a synchronized experience, we ensured that the first video seen by each participant was different. Doing so kept the different order situation realistic to sites that use playlists, as newcomers are likely to start off watching a different video than the other people in the group, but may find that as they continue watching, they see content that other people have recently watched (or are currently watching).



Three participants watch the same videos together.

Three participants watch the videos in different order.

Figure 3. Half the participants watched the same videos at the same time; the other half watched different videos at the same time.

2.1.4 Measures

We measured three kinds of enjoyment in this study: enjoyment of each video, enjoyment of the chat (for those participants with a chat feature), and enjoyment of the overall experience. We measured video enjoyment by having participants rate each video on a 5-point scale. To avoid difficulties in recall, participants rated each video immediately after they watched it. We measured chat enjoyment by having participants rate three statements about the chat on a 5-point Likert scale. These questions were averaged to form a scale of chat enjoyment with good reliability (Cronbach's $\alpha = .79$). We measured enjoyment of the overall experience by asking participants to rate the experience of participating in the study on a 7-point item, where answers could range from "very boring" to "very fun". Table 1 summarizes our questionnaire measures.

We hypothesized that audio chat would be more engaging and distracting than text chat. We measured engagement by asking participants how much attention they paid to the chat. We measured distraction in two ways: first, by asking participants how distracted they were by the chat; and second, by using a memory measure. Memory measures are less influenced by whether distraction is verbal or visual (cf. [9]). We asked one multiple-choice question about each of the 8 videos, for a maximum correct memory score of 8. For videos with dialogue, the questions asked about what performers in the videos had said. For videos without dialogue, the questions asked about what performers had done or the locations shown in the video.

To compare the chat between our different conditions, we logged the chats of our participants. We recorded and transcribed the conversations of participants who used audio chat. For these groups, speech was transcribed such that one thought or phrase corresponded to one line in the transcript. For example, when two speakers alternated in speaking, each alternating turn was a separate line in the transcription. When one person spoke, paused for a moment, and then spoke again, the pause was considered to be the beginning of a new conversational turn, and was placed on a separate line. We used two seconds as a rough guideline for the length of these pauses, but also considered whether the content after the pause was related to what was previously said. For example, "Are they fishing for something? (*pause*) It looks like they have nets" was kept together, because the statement gives the reason for the question.

2.2 Results

We analyzed the data using correlations and analyses of variance (ANOVA). In our ANOVA, the model compared the independent variables of chat media (no chat, text chat, audio chat, both text & audio) as a between groups variable, dialogue presence as a within subjects factor (no dialogue vs. dialogue), video synchrony (same video order vs. different order) as a between groups variable, and their interactions. Because the data are from individuals who watched videos together in a group, we used a mixed model that accounted for group as a random factor.

Table 1. Questionnaire measures used in the study.

Measure	# Items	Items
Chat enjoyment	3-item, 5-point Likert scale ($\alpha = .79$)	*Example item:* "I enjoyed talking with the people in my group while watching the videos."
Video enjoyment	8 items (one per video)	"After each video finishes, please circle your rating below (5 is highest),"
Overall enjoyment	1 item, 7-point item	"How would you rate the experience of participating in this study?"
Chat engagement	1 item, 7-point item	"How much attention did you pay to the chat?"
Distraction (self report)	1 item, 7-point item	"How distracted were you by the chat during the videos?"
Distraction (memory)	8 items (one per video)	*Example item:* In Paddy the Pelican, why did the boat not start? A) Out of gas. B) Broken rudder. C) Filled with water. D) Missing oars. E) I do not recall
Media comfort	2-item, 5-point Likert scale ($\alpha = .71$)	*Example item:* "I felt comfortable *{typing, speaking}* while watching the videos"
Media preference	1 item	"Which type of chat would you have preferred to use?"

2.2.1 Chat amount and content

We controlled what videos people saw but not what they said in the chat conditions. Therefore, to understand the effects of chat on enjoyment and distraction, we first examine the amount and content of chat. To do this, we iteratively developed a coding scheme for the chat in this study based upon the scheme used in [21]. We used the line of chat as our unit of analysis. The coding scheme used in this study is summarized in Table 2. Our entire corpus contained 10,812 lines of chat, and a subset of 869 lines (8%) was used in a reliability analysis. Two independent coders coded this subset of chat and achieved a Cohen's Kappa of .71. This is an adequate level of reliability for our analysis [12].

Table 2. Chat coding scheme.

Category	% chat (total)	% chat (same order)	% chat (diff. order)
Discuss video	36.1	45.3	25.0
Discuss study	17.8	15.6	20.4
Laughter	16.9	17.5	16.2
Rate video	12.0	10.5	13.8
Personal	9.9	7.5	12.8
Coordination	5.4	1.4	10.3
Greetings & partings	1.4	1.3	1.4
Unintelligible	0.5	0.9	0.1

In comparing the amount of chat produced by different groups, we use word counts instead of line counts because they are a more accurate measure of how much each group chatted.

Chat media had an effect on the amount of chat. Participants with text chat produced 1,020 words (SD = 379 words), and participants with audio chat produced 2,202 words (SD = 1,379 words), showing that people chat audibly about twice as fast as they type. This difference is significant ($F [1,30] = 8.15$, $p < .01$). Groups with both text and audio chat features mainly used audio chat; they typed an average of 401 words (SD = 370 words) and spoke an average of 1,405 words (SD = 729 words), and this difference was significant ($F [1,20] = 15.5$, $p < .001$).

Chat media had an effect on the distribution of conversational topics ($\chi^2(14) = 160.2$, $p < .001$). One difference is that groups with text chatted more about the videos (43%) than groups with audio (37%) or both text and audio (37%). Another difference is that groups with audio or both text and audio laughed a little more than groups with text (~13% vs. 7%).

Video order also had an effect on the distribution of conversational topics ($\chi^2(7) = 974.2$, $p < .001$), as shown in Table 2. One key difference is that roughly 10% of the chat in different order groups was about coordination. This was hardly present in same order chats. Also, more of the chat was focused on the videos themselves in same order groups (45%) than in different order groups (25%). Different order groups had more chat focused on personal topics (12.8%) than same order groups (7.5%).

Finally, there was a minor difference in the distribution of conversational topics between videos with dialogue and videos without dialogue. For videos with dialogue, 21% of the chat was laughter, and for videos without dialogue, 15% of the chat was laughter. Since jokes are usually spoken rather than visualized, it is quite possible that the videos with dialogue were funnier.

2.2.2 Enjoyment

Overall, participants enjoyed participating in the study (M = 5.0 [SD = 1.3] on the 7-point scale). They rated videos a little above midway on the 5-point scale (M = 2.9 [SD = .5]. They rated the chat feature slightly above average on the 7–point scale (M = 3.8 [SD = .9]). Participants' overall enjoyment of the study was most closely correlated with their ratings of the videos ($r = .49$, $p < .001$), and (for those who chatted) it was not related to their enjoyment of chat ($r = .05$). This finding suggests that the experience of chat is independent of the experience of watching videos.

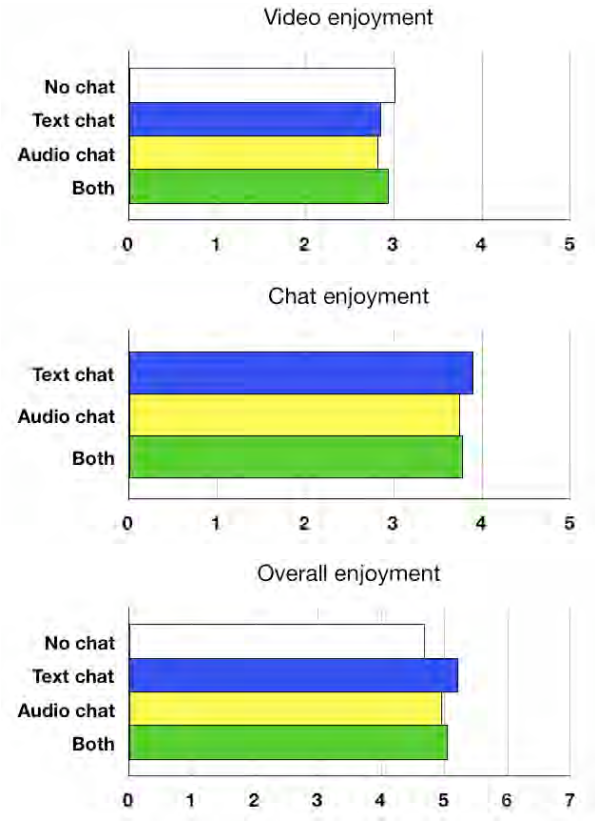


Figure 4. Enjoyment by chat media.

We predicted that audio chat would be more engaging and enjoyable than text chat. We did not find this to be the case. We tested "engagement" by asking participants to rate how much attention they paid to the chat. Groups with text chat reported highest attention to the chat, M = 5.7 (SD = 1.2) whereas those with audio chat reported a mean chat attention of 5.0 (SD = 1.5) and groups with both text and audio reported a mean chat attention of 4.6 (SD = 1.5). There was a marginally significant difference overall among the three chat conditions ($F [2,30] = 3.0$, $p = .06$). Individual contrasts show that groups with text were more attentive to the chat than groups with text and audio ($F [1,30] = 5.92$, $p = .02$). Groups with audio alone were not significantly different from either group. Attention did not significantly differ among groups watching in the same order or in a different order ($F [1,30] = .98$, p = n.s.).

As expected, engagement (attention to the chat) was associated with higher levels of enjoyment of the chat ($r = .41$, $p < .001$). A comparison of enjoyment across the chat media conditions is shown in Figure 4. The data shown suggests a trend that text chat was the most enjoyable form of chat, that not having chat increased video enjoyment, and that having chat increased overall enjoyment. These effects are not significant. Thus, overall, we conclude that audio chat and text chat (or both) were equally enjoyed in our study, and generally did not change the video experience. However, usability remains important. Our measure of media comfort using the specific chat media (either text or

Table 3. Correlations among constructs.
***p < .05, †p < .01, ‡p < .001**

	1	2	3	4	5	6
1. Overall enjoyment	1.0					
2. Video enjoyment	.49‡	1.0				
3. Chat enjoyment	.05	-.06	1.0			
4. Distraction (memory)	.10	-.02	-.12	1.0		
5. Distraction (self report)	.03	-.03	-.27†	.21*	1.0	
6. Chat engagement	.09	-.04	.41‡	.03	-.09	1.0
7. Media comfort	.13	-.16	.28†	-.14	-.05	.17

audio) was associated with higher levels of enjoyment of the chat (r = .28, p < .01).

2.2.3 Distraction

We used two measures of distraction in this study. They were positively correlated and thus have modest concurrent validity (r = .21, p < .05; Table 3).

The first measure was a subjective self-report measure; participants were asked how distracted they felt on a 7-point scale. Ratings of distraction generally fell in the middle of the scale (M = 3.8 [SD = 1.7]). In comparing the different chat media conditions, we did not find any significant differences among groups with text, groups with audio or groups with both text and audio (text: M = 3.6 [SD = 1.6], audio: M = 4.2 [SD = 1.9], both: M = 3.7 [SD = 1.6]). We conclude from this analysis that if participants were differentially distracted, they were unaware of it. Feelings of distraction did not correlate with how much participants chatted (r = .09, p = n.s.), which also suggests that people were unaware of distraction when it occurred.

To evaluate how distracted participants actually were, we asked eight multiple-choice questions about the video content. Overall, participants got 5.8 questions correct (SD = 1.4 questions); 20 participants (14%) got perfect scores.

There was no main effect difference between the scores of groups with text chat, audio chat, or both text and audio chat. Groups with text chat answered an average of 5.4 questions (SD = 1.6 questions) correctly, groups with audio answered an average of 5.8 questions (SD = 1.3 questions) correctly, and groups with both text and audio answered an average of 5.4 questions (SD = 1.3 questions) correctly.

Comparing groups with a chat feature and groups without a chat feature, groups without chat answered an average of 6.7 questions (SD = 1.2 questions) correctly and groups with chat answered an average of 5.5 questions (SD = 1.4 questions) correctly. The contrast (from the overall ANOVA) between groups without chat and groups with chat shows that this is a significant difference (F [1,40] = 16.8, p < .001). Also, of those with a perfect memory score, 55% were in the no chat condition. Participants' memory scores were negatively (but insignificantly) correlated with how much they chatted (r = -.11).

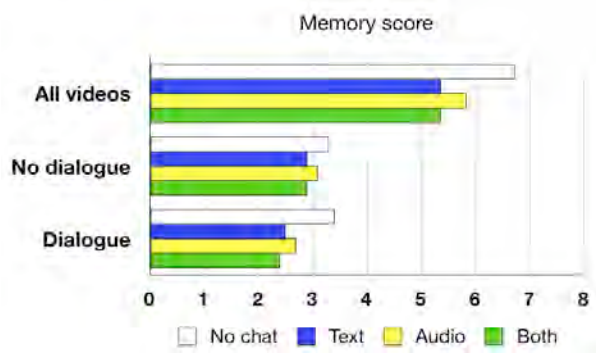


Figure 5. Memory scores for all 8 videos, 4 videos with no dialogue and 4 videos with dialogue.

These findings show that despite their lack of awareness of distraction, those who chatted (whether using audio or text chat) were more distracted than those who did not have a chat feature. Thus, chatting did detract from memory performance, although there was little correlation between actual distraction and overall enjoyment (r = .10) or enjoyment of the videos (r = -.02).

We next address the question of how distraction differed based on the chat media used and the types of videos being watched (dialogue vs. no dialogue). Recall that the presence of a chat feature caused a significant reduction in memory scores. We now examine this finding in more detail, comparing memory for videos with dialogue and videos with no dialogue.

For videos with dialogue, the questions on the multiple-choice memory test asked about events or scenes in the video. For videos without dialogue, the questions on the memory test asked about what participants heard. Memory scores on no dialogue videos had a maximum value of 4 and memory scores on dialogue videos had a maximum value of 4.

Participants' memory scores are shown in Figure 5. As a baseline, participants without chat did well; they had a mean score in the no dialogue condition of 3.3 (SD = .8) and a mean score in the dialogue condition of 3.4 (SD = .9). Participants with text chat had a mean score in the no dialogue condition of 2.9 (SD = 1.0) and a mean score in the dialogue condition of 2.5 (SD = 1.0). Participants with audio chat had a mean score in the no dialogue condition of 3.1 (SD = .8) and a mean score in the dialogue condition of 2.7 (.9). Participants with both text and audio chat had a mean score in the no dialogue condition of 2.9 (SD = .8) and a mean score in the dialogue condition of 2.4 (.8).

For videos with no dialogue, participants with chat tended to remember less than participants without chat (F [1,40] = 2.8, p = .10). For videos with dialogue, the negative effect of chat was greater. Participants with chat remembered less about videos with dialogue than participants without chat (F [1,40] = 18.4, p < .001).

Comparing between videos, participants with audio chat or both text and audio remembered less of videos with dialogue than videos without dialogue (audio: F [1,272] = 4.51, p = .03; both: F [1,272] = 5.09, p = .02). The difference with text chat was a trend

in the same direction but only of marginal significance (F [1,272] = 2.98, p = .08).

Comparing between groups with text, audio or both text and audio, there was no difference in how much they remembered from videos without dialogue (F [2,30] = .71, p = n.s.). There was also no difference in how much they remembered from videos with dialogue (F [2,30] = .54, p = n.s.).

In sum, we conclude that each type of chat was equally distracting, and especially when watching videos with dialogue.

2.2.4 Audio chat in a playlist model

Our third research question predicts that audio chat will be particularly distracting when viewers lack a shared media context, as in the playlist model, because viewers will be unable to coordinate their speaking to coincide with "quiet" times in the video. This prediction was partly supported by our data (see Figure 6). Different order audio groups tended to remember less than same order audio groups (F [1,40] = 2.3, p = .13). This difference depended on the types of videos being watched. For videos with dialogue, there was no difference between the same order and different order groups. For videos without dialogue, viewers were more distracted when they watched those videos in a different order than when they watched them in the same order (F [1,40] = 4.4, p = .04). This suggests that same order groups were able to better coordinate their speaking in the absence of interfering dialogue.

2.2.5 Media comfort & preferences

We asked participants about their comfort using the different chat media, as well as their preferences for them. Groups with text chat reported a mean comfort score of 3.7 (SD = .8), and groups with audio chat reported a mean comfort score of 3.6 (SD = .9) on a 5-point scale. Groups with both text and audio were asked the questions on the media comfort scale twice, once for each media type. They reported a mean text comfort of 3.7 (SD = .8) and a mean audio comfort of 3.6 (SD = .9).

To gauge peoples' preferences for each media type, we asked participants at the end of the study which medium they would have preferred to use, from all available options: no chat, text chat, audio chat, or both text and audio.

Overall, participants reported preferring having chat (86%) to not having chat (14%). However, the most striking difference is between groups with text and groups with audio. The responses for groups with text chat are shown in Figure 7a, and the responses for groups with audio chat are shown in Figure 7b. For participants with text chat, 56% reported preferring it to the other options. A little more than a third (36%) felt they would have wanted the option of audio (by responding audio or both). However, for participants with audio chat, almost 70% reported wanting audio in some form (by responding audio or both), and a little less than a third (30%) did not prefer it. These results are even more pronounced for groups with both text and audio: only 3% would prefer not to chat, whereas 28% would want just text, 22% would want just audio, and 47% would want both text and audio. Thus, there was a strong preference for audio chat; people with audio preferred it to text chat or no chat.

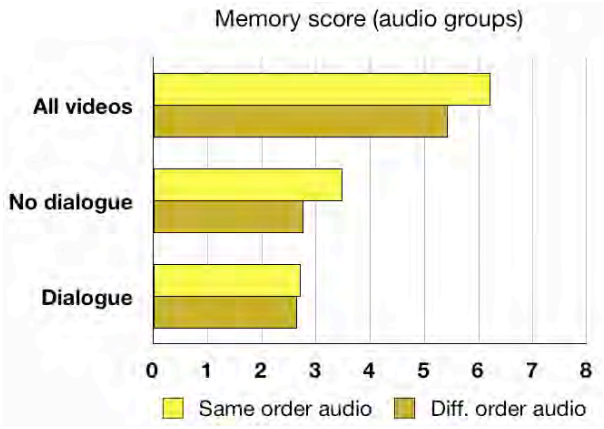
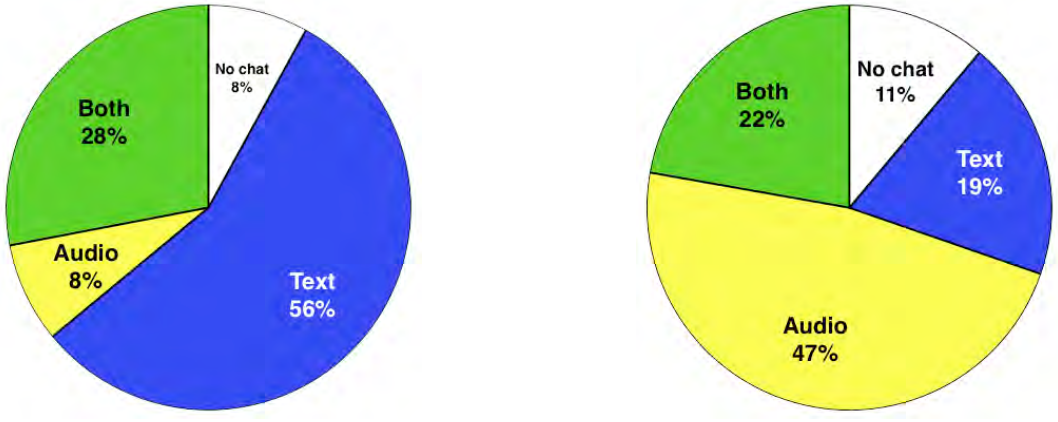


Figure 6. Memory scores for audio groups.

However, these results differ when we asked participants to speculate on the experience of chatting with strangers. In this instance, participants expressed a strong desire to use text chat (62%) over audio chat (22%) if they watched with strangers (the last 16% reported not wanting chat). Even among those who had audio chat in the study, 53% reported wanting to use text chat with strangers. Several reasons were given for why participants preferred text chat with strangers, including "I'm shy around strangers" (P123), "[It's] less intimidating" (P7) and "Audio chat is more suited with friends. With strangers there can be some



(a) (b)

Figure 7. Chat media preferences for (a) text groups and (b) audio groups.

awkwardness initially" (P141).

To gauge people's sensitivity to how the chat media interacted with the content being watched, we asked them if they felt that different chat media were suited to different types of videos. Overall, about half of our participants felt it did, and of those who had both text and audio, 75% felt chat media mattered. Participants were sensitive to the auditory distraction:

> *"If there is no talking in the video, chatting is ok, but when there was talking, the texting was more appropriate." (P3, both)*

> *"It's better to text if the movie relies heavily on dialog and is interesting." (P48, both)*

> *"I think using an audio chat would totally disrupt the attention given to watching a video." (P85, text)*

Other participants felt the asynchrony afforded by text chat could be beneficial.

> *"Text chat is always less distractive since I can answer it anytime I would like to, audio chat requires me to answer right away." (P18, both)*

Some felt that audio was advantageous because of its immediacy.

> *"Audio chat allows you to voice your immediate reaction." (P95, audio)*

> *"Audio chat helps to get the message quick and fast." (P57, both)*

Finally, for some participants, media simply didn't matter.

> *"To me it doesn't matter as long as you can talk to someone." (P21, text)*

3. DISCUSSION

Our results suggest that people do enjoy audio chat as much or more than they enjoy text chat. The groups of friends in our study who had both features, used them both and switched back and forth, but they favored audio by a wide margin (3 to 1). Both audio and text chat were equally distracting in the sense that they reduced memory for the videos. They were also more distracting when the videos contained dialogue. However, the effects were small, participants were not very aware of this distraction, and they did not enjoy the videos or the experience less as a result. Feelings of distraction, when they occurred, mainly detracted from the experience of chat rather than the video or overall experience.

We also tried a creative measure of engagement in the study. We gave participants a pretzel snack while watching, and speculated that participants would eat fewer pretzels when they were more engaged in the study. Groups with text chat ate significantly more pretzels than any other chat condition (29g vs. ~16g, $F[3,40] = 3.07$, $p = .03$). However, it is unclear if this meant that groups with text chat were less engaged; groups with audio may have simply been polite by not talking with their mouths full.

Participants who were in the no chat and text groups indicated that they would not like to use audio. Indeed, this was our own intuition before running this study; we too felt that audio would be too distracting while watching videos, and especially when watching videos with dialogue. On the contrary, those who had audio chat generally liked it. Once participants were exposed to using audio, they expressed a strong preference for it – about 70% of participants with audio (or both text and audio) preferred it.

There are several indications of the reason for this preference. First, audio chat was easy to use and allowed for more words to be exchanged. Second, audio chat seems to have been more social; groups with audio chat spent more of their chat laughing with each other than groups with text chat (20% vs. 10%). Laughing with others has been shown to increase one's perceptions of amusement [19]. Further, laughter in groups is often used as a signal to look for humorous material in the content being watched [6]. Although our questionnaire measures did not pick up differences in enjoyment between text and audio groups in this study, we speculate that groups that tried audio chat may have experienced stronger feelings of connectedness and presence with each other because of (or reflected in) their shared laughter.

The preference we saw for audio chat among those who used it is contrary to the result found in [17], where participants reported a greater preference for text chat (60% vs. 40%). One explanation for this is that their study surveyed classmates using text and audio chat while collaborating on a course project, whereas our study examined friends using text and audio chat in an entertainment experience. Our findings do mirror those of Geerts [7] who found that groups of friends and family liked using audio chat while watching television, even though it was distracting.

When our participants were asked about chatting with strangers, they reported a strong preference for text chat, even when they had used audio chat in the study. This highlights an important design decision for online video sites: who is the intended audience? If it is people who do not know each other, then text chat may be a more appropriate medium, as it supports larger groups and a higher degree of selective self-presentation [20]. If it is people who already have established relationships with each other, then audio chat may be more appropriate because of its higher degree of intimacy.

There were not many differences in the chats of participants who watched in a streaming vs. playlist model. Participants watching the videos in a different order (playlist model) chatted just as much as participants watching in the same order (streaming model). However, the distribution of their chat topics did differ between the two models. In line with our prediction, participants watching the videos in the same order were more on-topic (with respect to the videos) than participants watching videos in a different order. Different order groups focused more on other things they shared such as the study in which they were participating and personal topics. Further, about 10% of their chat was spent on coordination, asking their friends about what they were watching and telling their friends what they were watching.

Although one could argue that time spent on coordination could have been better spent on other topics, coordination may provide opportunities for viewers to elaborate about the videos they are watching and recommend videos to watch. Sites that use a playlist model have "rooms" or "channels" where people gather to watch together. Each channel contains a library of videos to watch. If these libraries contain an inordinate number of videos – e.g. hundreds or thousands – viewers may be unlikely to find other people in the channel who have watched the same content they are watching. Therefore, querying what videos other viewers are watching may serve to help introduce members to each other and bootstrap a social experience. Further, this querying can be automated, such that viewers receive recommendations for chat partners based on mutually having watched the same videos (albeit at different times).

3.1.1 Limitations

Our study has some threats to ecological validity inherent to laboratory work: the laboratory setting is not reflective of the real-world settings in which people watch online video; we cannot generalize to other type of videos (such as feature-length movies); and our population was generally comprised of young university students, so we cannot generalize to other populations. Indeed, prior work has shown that chatting while watching online video is not necessarily for everyone, and especially not for people who simply aren't interested in having social experiences online [21]. Further, our groups were comprised of several friends. Although preferences for chat media shifted from audio chat to text chat when participants were asked about chatting with strangers, their hesitation may be unwarranted. For example, using voice chat in video games has been shown to increase liking and trust, even among players with weak ties [23]. Therefore, researchers should examine the impact of audio chat on the social experience of watching online video with strangers.

Our study measured distraction in a naturalistic setting, where we allowed participants to chat as much as they wanted about any topics they liked. Although our results suggest that text and audio chat are equally distracting, further research should be conducted to measure distraction when the amount and topics of chat are more equivalent. Measures other than distraction are also important in studying chat with online video, and researchers should also consider immersion, self-consciousness, and social enjoyment (e.g. rating the chat). We speculate that the combination of chat with video might reduce social anxiety and self-consciousness because immersive videos take people's attention away from themselves.

Finally, video chat is an even richer medium than text or audio chat, and it may provide a more immersive social experience. However, video chat should come at the cost of additional distraction, as both the visual and auditory channels would be overloaded. Researchers should examine this issue as well.

4. DESIGN RECOMMENDATIONS

In our study we have made a few important comparisons to understand different types of online video experiences. Our results have several implications for the design of online video sites. First, we have demonstrated further evidence that the provision of a text chat feature is warranted as part of an online video experience. Even though we found evidence that chat is distracting (both objectively and subjectively), participants enjoyed using the chat feature, and the distraction did not diminish their video or overall enjoyment.

In comparing between text and audio chat, we found no substantial differences in distraction. The human factors and attention literature cited earlier suggests that audio chat may be more distracting than text chat. This is because when people attempt to listen to two simultaneous sources of audio (i.e. the audio from a video and the audio from people chatting), they experience a significant drop in their recall of the unattended audio channel [14]. This distraction did not discourage our participants. Further, when people were given audio chat, they overwhelmingly preferred using it with their friends. Therefore, we recommend that online video sites consider providing audio chat as a key feature, especially if the site is targeted to online communities whose members have offline friends. Some consideration does need to be made for video content and audience. For instance, audio chat may not be suitable for highly complex video content or informational or educational videos with heavy dialogue, since we found audio chat to be more distracting for content with dialogue. Further, text chat is capable of scaling to higher numbers of users, whereas audio chat can only support a small group – using audio chat with groups of more than a few others may become unmanageable.

5. CONCLUSION

Many online video sites are adding real-time chat as a key community feature. This chat is usually text-based, in which viewers type in comments to one another, and see them scroll by in a list box or as a 2D bubble above a graphical avatar. We are unaware of any online video site that allows viewers to chat with each other using audio. The reason is not a technological one, as audio chat has been successful in other types of online activities, such as gaming [23]. The results of this study suggest that audio can be successfully used for online video as well. Counter to our intuition, the groups of friends in our study used and enjoyed using audio chat.

We also compared between two models of online video sites: the streaming model, where all viewers watch the same content at the same time; and the playlist model, where viewers choose what they want to watch. Intuitively, viewers who watch different videos should experience difficulties in finding topics to chat about, because they lack the common ground of the videos. However, this was not seen in our study. Viewers watching the videos in a different order chatted just as much as viewers watching in the same order. Viewers watching in a different order also chatted less about the videos they were watching, but made up for this by chatted a little bit more about their own personal lives. They also kept each other up-to-date in terms of what they were watching. Both of these can be positive in the context of an online video community, where members bond with one another by recommending videos to each other and speaking about their own personal lives.

6. ACKNOWLEDGMENTS

We thank Jessica Wu and Tamir Sen for their assistance in running the laboratory study and transcribing audio logs. This work was supported by National Science Foundation grants IIS-0325049, CNS-0520187, CNS-0085920 and CNS-0435382.

7. REFERENCES

[1] Abreu, J., Almeida, P., and Branco, V. 2001. 2BeOn: interactive television supporting interpersonal communication. In *Proceedings of the Sixth Eurographics Workshop on Multimedia 2001* (Manchester, UK, September 8-9, 2001). J. A. Jorge, N. Correia, H. Jones, and M. B. Kamegai, Eds. Springer-Verlag New York, New York, NY, 199-208.

[2] Benford, S., Greenhalgh, C., Craven, M., Walker, G., Regan, T., Morphett, J., & Wyver, J. 2000. Inhabited television: broadcasting interaction from within collaborative virtual environments. In *ACM Transactions on Computer-Human Interaction,* 7 (4), 510-547.

[3] Coppens, T., Trappeniers, L., & Godon, M. (2004). AmigoTV: towards a social TV experience. In J. Masthoff, R. Griffiths, & L. Pemberton (Eds.), *Proceedings from the Second European Conference on Interactive Television "Enhancing the experience"*, University of Brighton.

[4] Daft, R. L., & Lengel, R. H. 1986. Organizational information requirements, media richness and structural design. *Management Science,* 32 (5), 554-571.

[5] Dennis, A. R., & Kinney, S. T. 1998. Testing media richness theory in the new media: the effects of cues, feedback, and task equivocality. *Information Systems Research,* 9 (3), 256-274.

[6] Fuller, R. G. C., & Sheehy-Skeffington, A. 1974. Effects of group laughter on responses to humourous material, a replication and extension. *Psychological Reports,* 35 (1), 531-534.

[7] Geerts, D. 2006. Comparing voice chat and text chat in a communication tool for interactive television. In *Proceedings of NordiCHI 2006.* ACM, New York, NY, 461-464.

[8] Harboe, G., Metcalf, C. J., Bentley, F., Tullio, J., Massey, N., and Romano, G. 2008. Ambient social tv: drawing people into a shared experience. In *Proceedings of SIGCHI 2008.* ACM, New York, NY, 1-10.

[9] Harms, L., & Patten, C. 2003. Peripheral detection as a measure of driver distraction. A study of memory-based versus systems-based navigation in a built-up area. *Transportation Research Part F: Traffic Psychology and Behaviour,* 6 (1), 12-36.

[10] Jeong, S-H., & Fishbein, M. 2007. Predictors of multitasking with media: media factors and audience factors. *Media Psychology,* 10, 364-384.

[11] Keele, S. W. 1973. *Attention and Human Performance.* Pacific Palisades, CA: Goodyear.

[12] Landis, J. R., & Koch, G. G. 1977. The measurement of observer agreement for categorical data. *Biometrics,* 33 (1), 159-174.

[13] Luyten, K., Thys, K., Huypens, S., and Coninx, K. 2006. Telebuddies: social stitching with interactive television. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems.* ACM, New York, NY, 1049-1054.

[14] Moray, N. 1969. *Listening and Attention.* Baltimore, MD: Putnam.

[15] Norman, D. A. 1968. Toward a theory of memory and attention. *Psychological Review,* 75 (6), 522-536.

[16] Regan, T. and Todd, I. 2004. Media center buddies: instant messaging around a media center. In *Proceedings NordiCHI 2004.* ACM, New York, NY, 141-144.

[17] Scholl, J., McCarthy, J., and Harr, R. 2006. A comparison of chat and audio in media rich environments. In *Proceedings of CSCW 2006.* ACM, New York, NY, 323-332.

[18] Shamma, D. A., Bastea-Forte, M., Joubert, N., & Liu, Y. 2008. Enhancing online personal connections through the synchronized sharing of online video. In *Proceedings of SIGCHI 2008.* ACM, New York, NY, 2931-2936.

[19] Smyth, M. M., & Fuller, R. G. C. 1972. Effects of group laughter on responses to humorous material. *Psychological Reports,* 30 (1), 132-134.

[20] Walther, J. B. 2007. Selective self-presentation in computer-mediated communication: hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior*, 23 (5), 2538-2557.

[21] Weisz, J. D., Kiesler, S., Zhang, H., Ren, Y., Kraut, R. E., and Konstan, J. A. 2007. Watching together: integrating text chat with video. In *Proceedings of SIGCHI 2007.* ACM, New York, NY, 877-886.

[22] Wickens, C. D. 2002. Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science,* 3 (2), 159-177.

[23] Williams, D., Caplan, S., Xiong, L. 2007. Can you hear me now? The impact of voice in an online gaming community. *Human Communication Research*, 33 (4), 427-449.