Talk Space: Developing the Online Salon

by

Karthik Palaniappan

Supervised by Alan Borning

A senior thesis submitted in partial fulfillment of
the requirements for the degree of

Bachelor of Science
With Departmental Honors

Computer Science & Engineering

University of Washington

March 2015

Presentation of work given on December 11th, 2014

Thesis and presentation approved by Alan Borning

Date March 15th, 2015
1. Introduction

I was inspired to pursue a better space for online discussion because of my own experiences with discussions in-person and online. Rather than formal, structured discussions, I prefer unstructured, small group discussions. However, I have never been able to reproduce that same feeling online, especially in text based forums. These discussions lack the same “intimacy” as in-person discussions, especially when conversations include more than a few people. My main goal through this project was scaling online discussions while preserving the quality and intimacy of in-person discussion.

Very early in the project, I stumbled across the “Red Chair Salon”, and I was reminded of something that shared its name: the “salons” of Enlightenment Europe. Salons were a gathering sponsored by a patron to increase the knowledge of the participants through conversation. They were not only used for talking about politics, but also culture and art. Though the old salons were reserved for the bourgeois, Jurgen Habermas explores the “public sphere” created by salon-type discussions as a way for private individuals to influence government. However, he argues that currently, the public sphere is in decline while institutions and organizations, such as the news media, control the debate (Habermas 1962). Instead, Roger Soder calls for a “thoughtful public”—one that actively discusses and shapes public policy (Soder 2001). The concept of a salon struck me as a unifying theme for my desire for high quality discussions. The online salon is a 21st century platform for everybody—of all classes, races, and genders—to participate and influence the public sphere.

In designing an online salon, I focused on the notion of creating “quality scaled discussions”. Though there are already online tools for discussion among large numbers of people (e.g. chat rooms, Twitter, Disqs), I will review research on computer mediated communication that shows that few produce deep, intimate dialogues. Instead, these tools present a relatively unorganized listing of opinions. We explore how to scale discussions—increase the number of participants—while preserving the quality found in small group discussions.

This paper explores the essentials of quality, scaled discussions online, arguing that dynamic, participant-structured discussions between small groups of people create an excellent platform for online salons. I begin by walking through the four major design phases we explored. The lessons learned from these designs and prototypes culminate in Talk Space, the current implementation of the online salon.

Then, we examine the three overarching “qualities” of the salon that we learned through these phases. First, quality discussions rely on a degree of what Erickson and Kellogg call “social translucence”—the feeling that you are talking to a real person and can perceive social cues such as body language (Erickson and Kellogg 2000). In real life, we use cues such as nodding, fidgeting, or yawning to shape what we say, but online communication systems are generally “opaque” to these cues. The authors argue that adding elements of translucence helps avoid “trolling” by enabling people to leverage social norms found in face-to-face communication. Second, there should be a way to break conversations up into “groups”, and users should be allowed to dynamically create, join, and leave these groups. Some situations call for intimate conversations between very few people, while others call for a larger group of less involved participants. The salon should be flexible enough to support the variance in both group size and
discussion style required in these situations. Third, it should be easy for users to move between these smaller groups. Fluid movement between smaller groups allows participants to “choose who to talk to”, and shape conversations to fit their needs.

Talk Space is one embodiment of these three qualities. It is based around the concept of groups—it is a multi-room video chat in which participants can see, join, and leave different video chats. It presents a simple, unstructured UI that allows participants not only to decide which groups to join, but also negotiate group size and conversation topics, similar to a seminar discussion in real life. Both the video nature of Talk Space discussions and smaller features (such as the ability to eavesdrop on other groups) help increase the intimacy and translucence of online discussions.

Finally, I discuss how Talk Space is the bare minimum implementation of a salon, and can be extended with future protocols to support various types of virtual discussions.

2. Designing the Online Salon

In this section, I examine the various designs and ideas that eventually converged into the small, dynamic group discussions of Talk Space. While I discuss our work as four separate design sections, they actually reflect a continuum of ideas that build upon one another.

2.1. Starting point: video chats

One of the first design decisions I made was using video chats as a basis for the online salon. Salons rely on participants having a high degree of social presence: intimate, personal discussions where participants feel like they are interacting with other human beings (Short, Williams and Christie 1976). Travis Kriplean, et al. explore this problem when designing a text-based interface, Reflect for active web discussions:

> Active listening is integral to communication. Listeners must provide evidence to speakers to show that they are being attentive and understand. Listeners operate a backchannel where they nod, say “uh huh”, tilt their heads, finish sentences, and preface responses with a simple restatement of what the speaker said. These explicit acts of listening helps speakers “debug” their messages, as well as provide evidence that they are being recognized and heard. It also helps listeners demonstrate good faith as conversation partners; partners who provide more feedback through the backchannel are perceived as more patient, polite, and attentive. This process of trading off speaking and listening to establish mutual understanding is called grounding (Kriplean, et al. 2012).

Text-based applications offer very low social presence, but video chats preserve most of the visual and audio cues of discussions. This has the side effect that while video chatting, participants are more encouraged to preserve “face”—in other words, act according to social norms. On the other hand, text chatters can anonymously attack a name without having to directly attack the person to their face (Freiermuth 2011), or may resort to unusually strong negative or positive language to compensate for the lack of intimacy (Kiesler, Siegel and McGuire 1984).

However, video chats do not perfectly preserve these backchannels. Audio and visual cues are directional in real life, but are non-directional (all directly in front of you) with video chats.
Thus, the important backchannels of eye-gazing and head-turning are lost, and “who is speaking” changes far less (O’Conaill, Whittaker and Wilbur 1993).

In addition, without high bandwidth connections, video chats are plagued with lag and are often low quality, which further prevents cues such as nodding or interrupting (O’Conaill, Whittaker and Wilbur 1993). While these are important issues to address, we do not expect many of them to be a problem in the future as bandwidth and computing power are constantly increasing. Talk Space, for example, uses peer-to-peer video chats, meaning that if participants are close by it is faster to send their video streams directly to others rather than going through a server. When testing with fast Wi-Fi on the University of Washington campus, we noticed that audio/video lag was virtually undetectable.

Without a three dimensional space, video chats will never be as rich as real life, but our design decisions below attempt to add new cues (such as a text channel) on top of video chats.

2.2 Adding a text channel to video chats: Sliders

From personal experience, my intuition was that the most productive discussions occurred in small group discussions of 5-6 people or fewer. In fact, this intuition is grounded in social psychology. Hare notes that “although the size of the ‘natural’ group varies with the age and other social characteristics of the population…Observers find it easier to identify leaders in discussion groups of about six members than in either smaller or larger groups” (Hare 1981).

I initially focused on increasing the approximate “natural” group size with new channels of communication. I took inspiration from existing discussion types and online tools, and explored new forms of online discussions. Our goal was to maintain the intimacy of small group discussions, even while adding more people to a discussion.

Snapshots of opinions in text

Our first intuition was to explore a way for participants to communicate their opinions via text. Text is advantageous because it is faster; everybody can write down their opinions at the same time, as opposed to waiting for everybody to “go around the room” stating their opinions. Reading is also much faster than listening (Jones, et al. 2005).

In larger face-to-face discussions, participants holding minority opinions can feel uncomfortable expressing their opinions in the larger group, but are willing to speak out more over text chat (McLeod, et al. 1997). In addition, non-native (English) speakers often feel more comfortable expressing their views in a text format (Wang, et al. 2014). Thus, we saw a mix of elements from face-to-face and text chat as complementary.

Text also preserves a record of the conversation. During a discussion, participants would have a clear history of what they talked about, and think about where they might want to go.

Sliders: non-linear text channels

Our attempt to introduce a text channel to video imposed more structure than the unorganized listing of thoughts afforded by comment boards and chat rooms. We organize participant opinions on a sliding scale (a “slider”) in real time, drawing inspiration from ConsiderIt, a powerful platform for organizing opinions. ConsiderIt presents a view where you can see
everybody’s opinions on a sliding scale from “pro” to “con” on a single issue. This helps organize opinions both by issue, and where the person stands. Still, “pro” to “con” scales are not great representations because opinions are more nuanced—what does “60% pro 40% con” mean for different people? However, we expected that the conversations that arise from seeing where other participants fall on the scale would uncover the nuances in their opinions.

We observed that discussions often focus around a limited number of subtopics—“discussion points”—at which people stop and explore. Thus, we envisioned a “slider bar” for each discussion point. People would place themselves at a certain point on the scale, and express a summary of their opinion on the subtopic.

We decided to allow anyone to add a discussion point, and expected that a point would appear after it was clear that the conversation had shifted to that specific point. The slider opinions would provide a basis for the conversation to continue. Then, we expected that people would change their opinion and location on the slider as the conversation progressed.

Ascii Prototype
We made a simple prototype using an ascii slider on Google Docs, and used Google Hangouts for the video chat. This prototype did not allow anyone to specify their opinions on the slider, but did allow people to place themselves on the sliding scale using their first initial. Though it is not shown here, the left side of the scale represented “oppose”, the right side represented “support”, and the middle meant neutral or a mix of both. Two friends and I discussed immigration and immigration reform using this prototype.

Figure 1: Using the ascii slider prototype

The prototype session confirmed our expectations. We discussed immigration reform for a few minutes, and the discussion started focusing around the “American” identity (or lack of one). I created the point, as well as the ascii slider, and invited everyone to place their initial on the slider. At first, S was further to the left (con), and R and K were further apart. Though it looked like there were two different camps, the participants started to discuss why they placed themselves at certain places—and the reasons were surprising. Eventually, S moved closer to the center, while R and K moved closer together. The discussion helped showcase the nuanced views of all participants, something that may have been hard in pure text.

Interestingly, we started to find consensus more when we collectively started to play with the wording of the question. For example, after some debating, we raised the question, “how do we delineate sense of obligation from identity”, and actually all agreed that we do not all have the
same sense of obligation though we may believe in the same American identity. In discussions, how the problem is defined changes the solutions proposed. We find it important to allow a democratic way of negotiating and defining “points”.

**Functional prototype**

Using these two important concepts—the ability to write and change your opinion, as well as negotiate the “points”—we built a fully functioning prototype.

Any user could create a “point”, and it would appear on the page with its title (“This is an example point”), and a slider bar. Each user had a slider as well as a text bubble to update their opinions. They would also be able to see everybody else’s opinions below their bar by hovering over the corresponding slider (e.g. “Charlie”). The real-time nature of this application is important to see how people’s opinions change as the discussion progresses.

The other important piece is the ability to negotiate the wording of points. We decided that rather than allowing anyone to change the wording of points, and potentially change the intent of the person who posted the point, we would only allow the person who posted the point to change it. However, we allow other people to “request edits” to the point creator, and the creator will be able to “view edit requests”. Figure 2 shows a sample point from the perspective of the creator.

**Lessons trying out the real version**

I hosted discussions of five to seven people with the sliders, but sliders actually hindered the discussion. One major area of confusion was the notion of a “point”. I originally envisioned points to represent the major subtopics of a discussion, but it is hard to tell what constitutes a major point until the discussion has moved on. By that time, it is no longer useful to create a point for it. People therefore, tended to use points as topics they wanted to discuss next. We would often see participants create several points before even completing one. Focusing on creating points constrains the discussion to those points too much, and interrupts the flow of the video chat.

On the other hand, users often became more comfortable simply video chatting, and did not utilize the sliders. When I nudged the participants to use the sliders, the conversation would abruptly stop, and people would record their opinions. Then, it was harder than expected to jump back into the discussion from where we left off. This awkwardness also generally prevented people from discussing opinions recorded on the sliders.
In general, people have a hard time focusing on two channels of information at once. The sliders could be used as a secondary tool—for example, a polling tool that pops up periodically through a conversation. It could also be used as a more organized comment board or discussion space when people do not have the ability to interact through video chat—for example, a presidential debate. However, sliders should not be the major focus of the user interface, and are thus not useful in creating the core of an online salon.

2.3 A small group feel in a large discussion: Fishbowls
While the sliders aimed to expand the quality of small group discussions into larger groups, we looked more carefully at existing discussion types to create the feel of a small group discussion in a larger group. We took inspiration from the “fishbowl”, in which there is only a small inner ring participating in a discussion at one time, and an outer ring of listeners.

Maria A. Prile, creator of the fishbowl, describes how fishbowl discussions can be used to create a seminar within a larger class. Her fishbowl has three to five students who have already prepared for the discussion sit in an inner circle. The rest of the class sits in an outer ring around this group. She also adds an additional two to three seats for rotating members of the audience to contribute. She finds that through the dynamism of the fishbowl, the class “develops a singular kind of esprit de corps which motivates participation and keeps the process afloat” (Prile 1993). In other words, she found a way to maintain the quality of seminar discussions while adding participants. The fishbowl is especially powerful because it leverages the collective knowledge of the group—everybody has the opportunity to participate and engage with the rest of the class.

Figure 3: The general fishbowl format. There is an inner circle of active participants, and the rest of the group listens. Members of the inner circle may have the opportunity to rotate with audience members. Source: Seattle Angel Conference, an organization bringing together entrepreneurs and investors.
While Prile was designing for a structured classroom discussion, our challenge was to reinterpret the fishbowl for a more open online space, where participants could freely leave and enter the discussion. We would not have the benefit of three to five previously prepared members to lead the fishbowl from the inner circle.

**Inspirations**

There are no online “fishbowls”, but there are technologies that employ similar principles of rotating “who gets to talk”. I examine the simplicity in involving users in panel discussions through video chats of the HuffPost system, as well as the democratic nature of turntable.fm.

![Figure 4: A panel including specially invited users on HuffPost Live. Notice that most users still can only interact through the comment forum on the bottom right.](image)

HuffPost Live is a live broadcast sponsored by the online newspaper, *The Huffington Post*. It features daily news broadcasts, as well as live panels that often include users via Google Hangouts or Skype. While the audience participation is refreshing, it does not fulfill the requirements for a democratic salon that we hope to create. The HuffPost Live staff selectively picks audience members for specific discussions through a screening process, including an up to thirty second audition tape. Most audience members only have the option to participate through the comment section. Instead, we envision online fishbowls to allow any users to rotate in the center.

Turntable.fm is a more democratic platform, but used for sharing music rather than for discussions. DJs on stage take turns playing music from their “DJ Queue” for the entire room, and anybody can become a DJ by clicking on a DJ spot when one opens up. Audience members can support a DJ by clicking the “awesome” button and making their avatar’s head bob up and down, or vote to skip them using the “lame” button. Thus, the DJ and song selection are fairly democratic.
Figure 5: A typical room in turntable.fm. On the stage are the DJs playing music. The audience effectively "votes" for the DJs using the "lame" and "awesome" buttons.

However, turntable.fm is often overwhelming and confusing to use because it presents too much sensory information. Similar to other Second Life platforms, it features myriad channels for communication, while lacking the “translucence” we strive to achieve. Wang, Anstadt, Goldman, and Lefaiver discuss how in their Second Life-based discussions, group members generally had a “feeling of unease within a new environment,” and noted the lack of verbal (audio) and body language for communication that make Second Life discussions difficult (Wang, et al. 2014). In a sense, Second Life provides exactly the wrong cues in a discussion. The feeling of “intimacy” in discussions comes from the ability to perceive the social cues in other human beings. Thus, Second Life is not an appropriate discussion platform for the online salon.

Designing the online fishbowl
Our design combines the simplicity of HuffPost Live and the more democratic voting system from Turntable.fm. We envisioned the inner ring of the fishbowl to be a video chat, with the rest of the audience mostly hidden from the UI. The inner ring would allow all three to five seats (depending on the size of the group) to be “up for rotation” with audience members. The main design challenge was how to “fairly” rotate audience members in and out of the inner ring. We decided on a set of democratically elected representatives from the group.

Turntable.fm allows people to DJ by simply clicking on an open “seat”. Rather than choosing inner-circle members by how quickly interested participants could pick an open seat, we decided to use a more clear visualization—a “talk queue”. When a space opens up (i.e. someone from the center rotates out) the first person on the queue would take their place.

To increase the diversity of opinions represented, and allow a more democratic system, we envisioned a channel for the audience to “elect” representative members. Participants who wished to join the center would add an up to 140 character summary of what they wanted to say to the “queue”, and everybody who wanted to hear what they had to say would “upvote” their
opinion. The queue would then be ordered by popularity of opinion. This channel would only be available to the audience members—not the inner circle—to avoid the “multiple channels” attention problem encountered in the sliders, and generally keep the inner circle discussion focused.

Interestingly, the main problems that came up were some of the same problems as job scheduling in operating systems. One could imagine a topic in which most of the room wants to hear one of two opinions, but two people hold a radically different view. To prevent starving out the minority opinions, we added “time” as a factor in ordering the queue. Thus, the queue would also take into account how long a participant has been waiting to speak. We decided to wait until after testing before deciding on a balance between upvotes and time.

Prile’s classroom fishbowl was meant to encourage participation, but if users are willing to attend an online fishbowl, they are likely to have strong opinions, and would be less willing to yield the floor. To prevent some participants from dominating the discussion, we created a time limit for how long you could stay in the center. We estimated three to five minutes, but again left the exact number to be determined by testing.

**Testing with the prototype**
To validate these ideas about an online fishbowl, I built a simple prototype “talk queue” that people could interact with on their computers or mobile phones. Currently, there are no video chat technology that fluidly allows viewers to rotate between being audience members and active participants. Google Hangouts on Air, for example, introduces so much visual overhead when a user joins that it disrupts the conversation. So, we decided to try a fishbowl in person, but to use the online queue to decide who would get to rotate into the center.
Figure 6: A simple prototype of the queue weighted by upvotes and time. The timer at the bottom signals how long until the person "on deck" will get to rotate into the center (and somebody already in the center is forced out).

I set up a fishbowl with the director of the University of Washington Honors program, James Clauss, who has led multiple philosophical discussions for honors students. For our fishbowl, we chose the topic of “ethical dilemmas”—for example, “is it morally acceptable to steal to feed one’s starving children?” We proposed to use half an hour to discuss using a traditional fishbowl, and the other half hour to try one using the online queue, which would enable us to compare the two conditions.

Unfortunately, only five participants showed up—we decided that we did not have a critical mass for a traditional fishbowl, but could still try a smaller version of the queueing prototype. We started by allowing everyone to add their opinions to the queue, and later created a distinction between the inner circle (those who could talk) and the outer circle. As expected, when we allowed participants to add opinions to the queue, there was immediately variance in opinion.

However, people were confused what an “upvote” meant. Democratically electing members implies that you would vote for things that match your opinion, but the intention of the upvote feature was to encourage people to upvote what they wanted to hear. We also noticed that people started upvoting all or most opinions to avoid hurting the feelings any participants. This is likely because we did it in person, so we did not see this is as a real issue.

Next, I allowed two people to be “in the center” at one time, and the others silently waited to move to the center. We found that the queueing system was more complicated than expected, and that users disliked the excessive structure of the fishbowl, especially having to wait for their
Developing the Online Salon

2.4 Finding the necessary components of an online salon

The failures from the excessive structure in both the sliders and fishbowls led us to revisit the essentials of an online salon. Rather than try to increase the number of participants in one conversation, we explored the idea of allowing discussions to break up into smaller subgroups.

Examining assumptions

Thus far, I assumed that keeping the entire group together at all times was important. I had reasoned that if the group were split into multiple subgroups, the individual subgroups would not have the benefit of the knowledge gathered in the discussions of other groups. However, the value of focused, small, structure-less discussions may outweigh the benefits of the collective knowledge.

I originally expected people to value the greater diversity of opinions achieved with a larger group. In reality, some want to talk to people who agree with them, and others want to talk to those who disagree with them—different people want to talk to different groups of people. We hypothesized that people would find “dynamic” discussions, where they can choose who to talk to, more valuable.

Finally, I assumed that users want to participate more—they want to actively engage rather than be passive audience members, and our experiences support my assumption. I designed the online fishbowl as an improved panel discussion, where the audience could rotate into the panel, rather than simply listening. However, we found that people want to be able to engage the entire time, and were easily bored on the outer-ring. In fact, high school teacher Jessica Young reports that many of her students on the outside of a fishbowl are easily bored, and become disruptive. Instead, she promotes the “small group scored discussion”—where students are broken up into groups of three or four peers—in which she sees far more participation as well as enjoyment (Young 2007).

Qualities of an online salon

The overarching goal of our online salon has been to develop an application that can support conversations of both desired intimacy and large numbers of people. At a social gathering, as groups grow larger, they split into subgroups, and thus retain intimacy. We envisioned similarly supporting multiple groups within one larger conversation “space”. I expand on the three “qualities” of the online that we discovered.

- An “ideal” group size varies depending on factors such as the participants, conversation topic, and culture (Kerr 1989). I personally prefer very small groups (2–3 people), especially to talk about very deep topics. On the other hand, if a politician were holding a Q&A session, he/she might have a significantly larger group but most members would
not speak. **An online salon needs to be dynamic enough to support groups of various sizes.**

- **An online salon must allow fluidity of movement between groups.** It should be easy to create a subgroup from a larger group, or join an existing conversation. In general, participants should be able to “choose who to talk to”. This is actually an advantage of an online salon—there are potentially many more people to talk to, and a better chance for you to find conversations that interest you.

- **Finally, we want to preserve “social translucence”—social cues for negotiating relationships.** While video is not a requirement for an online salon, from our experience, video chats provide the most rich social cues and feedback for intimate conversations. In addition, if we are to promote fluid switching between groups, we need translucence into the conversations of other groups as well. Being able to watch or eavesdrop on other conversations will provide a mechanism to judge whether you want to join it.

### 2.5 Dynamic discussions: Talk Space

These reflections on how to structure the online salon led us to compare our notion of a “salon” with parties—thinking about how in real life people naturally clump into and move between smaller groups in a larger social gathering. In this section, I look at the design decisions and prototypes that factored into Talk Space.

**Inspirations**

We imagined large social gatherings—“parties”—as new basis for the online salon. Our goal was to provide people with the social cues necessary to dynamically converse within and among smaller groups. Thus far, we have assumed video chat would provide enough “social translucence” to hold a discussion with a small group, but now, we would need to provide cues between groups.

Our first idea was to create a Second Life-like-world but using users’ actual video streams to represent them, instead of just avatars. We imagined a world in which you would be able to move around and talk to people within your “field of vision”. You would clearly hear people within your “field of vision”, and hear the ambient noise from other conversations, to preserve the feeling that you are part of a larger space. We later found that this approach has been implemented by an Australian startup, iSee.

![Figure 7: iSee, a platform that merged video chats and Second Life. Users can form groups and naturally talk to others as in real life by virtually moving and orienting their video stream in the three dimensional world.](image)
However, we wanted to avoid a major pitfall of Second Life—unnecessary sensory information that overloads the user. Similar to the fishbowls, it became hard to design the “rules” for a virtual world. What constitutes your “field of vision”? What happens if another user is in your field of vision (you are looking at them), but they are looking away from you? These are problems we also see in iSee. Though it mocks real life, it is difficult to orient yourself correctly, and navigate around other video streams, and through narrow passageways in the virtual world.

We needed a different perspective. Rather than trying to design online technologies that simulate real life, we needed to think about how the strengths of this alternate communication medium could be leveraged differently to achieve the principles of a salon. For example, we could design different social cues users can employ that might achieve similar goals as “nodding” or “eye contact”. Jim Hollan and Scott Stornetta call this design perspective striving to be “Beyond being there” (Hollan and Stornetta 1992).

With this new perspective, we revisited the fundamentals. Our first priority (and the core of Talk Space) was simplifying the process of “choosing who to talk to”. We realized that large conversations are often broken up into smaller side conversations or subgroups. The major challenge is finding a subgroup that interests you, and being able to switch conversations without being rude. In other words: creating, leaving, and joining these small groups. The features of Talk Space focus around this concept.

An inspiration for the design was “Unhangouts”, a project from MIT for “unconferences” online. Unconferences are meetings in which most conversations happen in separate breakout rooms, usually around different subtopics. In Unhangouts, these breakout rooms are represented by separate Google Hangouts.
Figure 8: Unhangouts allows you to break out into separate "breakout rooms", represented on the bottom left.

Though powerful for organized meetings, Unhangouts do not fit the democratic nature of the online salon. The topics for breakout rooms in Unhangouts are pre-decided, and each room can only hold a certain number of people. It does not easily facilitate the dynamic switching between groups that we see in real life. Instead, we envisioned a way for people to have simultaneous, traditional video chats within one space, and also preserve the cues necessary for dynamic groups.

**Original design of Talk Space**

A group is the fundamental unit of Talk Space, and is the term we give for one conversation between video chatters. A user would create a group by clicking on an empty space in the world, and specifying the subject of the conversation. Users would be able to fully see and hear their groups, and see the participants in other groups (but only faintly hear other groups’ audio). Users within a chat group would also be able to change the “subject” of the conversation as the discussion progresses.

Users could decide to join a video chat group using both the subject and visual cues from just watching the group. Unlike Unhangouts, there would be no maximum size to a group; instead, we expected group sizes to be governed by group dynamics. While it is possible to have thirty people in a conversation, we see that in real life, the conversation breaks up into multiple side conversations. We expected a similar phenomenon with dynamic video chats—people would break off when they started to feel like they would not have ample time to express their opinions.
Finally, we wanted to translate the concept of “ambient noise”—hearing other conversations in a room—to our online platform. Rather than isolating each video chat, we wanted to allow people to hear other conversations (virtually) near them at a lower volume. This mocks how in real life we can hear and eavesdrop on other group conversations while not part of them. To aid eavesdropping, we allowed users to mouse over other groups to hear them. We hoped this would provide a cue to users about whether they wanted to join a particular group.

**Prototypes of intuitions**

We decided to prototype and test the three pieces of Talk Space separately. First we tested the value of ambient audio and eavesdropping; then, we tested the idea of separate groups, and getting cues from being able to watch other video chats; and finally, we tested having “subjects” for groups.

To test the ambient audio and eavesdropping, I recorded multiple conversations between friends. I then placed six pictures on a blank webpage—one for each audio stream (a “group”). While users moused over a group, 80% of what they heard came from the group they were moused over, and the other 20% was the combination of all other streams. When not mousing over a group, a user would hear all streams equally.

We found that it was difficult to eavesdrop on other conversations, because without visual cues, there was no context to the discussion. It took easily 10-30 seconds for users to fully understand what a group was talking about, and feel comfortable enough to “join” it. With the same prototype playing “ambient noise” at 20% in the background, we tried having real-life discussions to mock being in a group. We found that the extra noise was irritating and disruptive to the conversation. On the other hand, users were fine with hearing their group at 20% while they eavesdropped on other groups at 80%. So, we decided to remove ambient noise while you were in a group, but play ambient noise while you were in “limbo” (not in any group). We tabled eavesdropping until we prototyped with a video stream.

The video prototype of groups was far more positive. It was similar to the audio prototype but using pre-recorded video chats from YouTube. The visual cues provided by video allowed users to narrow down possible groups to join—they either did not consider, or hardly considered the ones that appeared “boring” to them. In a sense, this phenomenon is akin to flipping through TV channels; users can easily pick which ones to skip over, and which ones to listen into. It was also easier to get the context of a discussion while eavesdropping on another group. This prototype validated the concept of dynamic groups, and the need for social cues and video.
Finally, we created a prototype for allowing participants to change the “subject” of a conversation as it progresses. I created a small webpage with blank rectangles representing groups. Each group had a subject displayed on top of it, and a textbox on the bottom for users to change the subject as the discussion progressed. To test, a few friends and I joined a Google hangout in one browser window, and opened this prototype in another. Interestingly, we noticed the prototype had similar problems to the sliders, as the title bar became a second channel of communication. The video chat felt natural, so even though the conversation topics changed, no users felt inclined to change the subject box. After I nudged people to change the subject, they started to abuse it—changing it to subjects they wanted to discuss, or even use as text chat (e.g. subjects of “lol” or “WTF”). We decided that since video and audio provide good enough cues to decide whether to join a group, the subject feature was not important. Thus, it was not included in the final Talk Space.
Groups are essentially the same as in the original prototype. The thick black (or blue) rectangles signify separate groups—a separate video chat. New groups are created by clicking on the “create” button the top left. Participants can join any group by simply clicking on it, or leave the space entirely by clicking the “leave” button. This “limbo” space (where you are not part of any group) is a way for users to eavesdrop into different conversations without committing to engaging with them. Though we want to allow full participation, it is important not to force participation; users can still passively watch active discussions.

As discussed with the original design, groups do not have a maximum size. Instead, the groups start large enough for two users, and then grow as more users join. Importantly, Talk Space guarantees that groups always have one space “vacant”, to make it clear that users can always join groups. The concept of growing groups also increases the number of video chats you can see on one screen, and reduces wasted whitespace.

Testing with Talk Space confirmed our intuitions that people would choose to move between groups at least partially because of the ability to see other groups. We frequently noticed that if we had two conversations in parallel—one in which participants looked bored, and another that featured more laughter and engagement—the users in the boring conversation would switch into the engaging conversation.

The concept of eavesdropping remains in Talk Space, but ambient noise is only heard while eavesdropping. Our original design of Talk Space allowed users to hear noise from nearby conversations in addition to their group to make it feel like they were part of a larger space. However, since we found that ambient noise was often disruptive and irritating both in real life, and online, we did not include it in the prototype. Ambient noise is only present when you are in “limbo” (not in a group) to provide the feel that you walked into a space with many separate conversations.
However, as we tested out our rules around eavesdropping and ambient noise, we noticed that users were confused about why volume levels constantly changed (they would unintentionally move their mouse to another group). To mitigate this problem, the UI shows the sources of your audio stream with shade of blue corresponding to how much audio you here from that group (or the standard black for groups you cannot hear). If you can hear one group exclusively, it appears dark blue. When you eavesdrop on another group, your group is a very light shade of blue, and the group you mouse over is an intermediate stage of blue. This mirrors the fact that you will hear your group’s audio at 20%, and the other group at 80%.

![Figure 11: An example of eavesdropping. I (top left) am listening to the conversation on the right. Notice that the groups are highlighted lighter or darker shades of blue depending on whether you can hear less or more audio from a group.](image)

During testing, we often noticed that users would use hand signals to get the attention of a user in a different group. Alternatively, they would use a different medium of communication (e.g. Facebook chat) to either privately message another user. We realized that in real life, certain visual cues need to be directed to specific people. For example, people often nod toward someone to indicate their desire to split into a separate conversation. However, because Talk Space completely separates groups, it is impossible to have “eye contact” or communicate directly with a user from a different group without joining their group.
To mitigate this problem, we designed the “ping” feature: a way to quickly send a message to a specific user. A participant clicks on another user’s video stream, and it pops up a modal that asks for their name and the message to send. Then, the name and message appears on the other user’s screen as a yellow toast notification. In the future, we plan to explore additional features that will add social translucence and improve the experience.

We still had not fully solved the problem of hand signals. Most video chat software (including Talk Space) has your video stream be a mirror image of you, but other participants’ video streams are flipped—preserving how you are used to seeing yourself in real life (in a mirror), and seeing others (straight on). Thus, when you make a hand signal to the left, others see you pointing to the right. This became a problem, as one participant would point toward another group, but his/her fellow group members would be confused about which direction he was pointing.
Figure 14: Atul (left) is actually pointing to his left, but I (right) see him pointing to the right. This presents a confusing situation.

Figure 15: Once we flip his video stream we both see him pointing to the left, and there is no ambiguity.

We decided this was a significant enough problem to either orient all video streams (including yours) to look like mirrors, or like you are looking straight on. After trying both, we ran into an almost traumatic experience—seeing yourself or others from a different perspective. As a Radiolab segment explains, facial characteristics such as hair-part influence how we view someone. A left hair-part may be seen as more masculine, while a right hair-part may be seen as more feminine (Walter and Nicholls 2013).

I informally surveyed a few friends, and we unanimously agreed that between the two options, flipping the other streams to be like mirrors (and keeping yours a mirror) is the least shocking. If we flip your video stream, when you move left, your video stream shows you moving right. This awkwardness outweighs any benefit of preserving the normal orientations of the remote participants.

3. Implementation of Talk Space
Talk Space is a web application written primarily in JavaScript. It employs the Redis data store to track the active users and the state of who is in which group. The backend uses node.js and socket.io to serve the HTML pages, push changes (e.g. a user joining a group) to all clients, and to forward private “ping” messages.
The video streams leverage the WebRTC peer-to-peer multimedia protocol. We capitalized on the developer API of vLine, a video chat startup which hides differences in browser implementations of WebRTC. Currently, only Chrome, Firefox, and Opera support WebRTC, limiting the usage of Talk Space. However, Internet Explorer, among others, plans to add support for WebRTC in future releases.

The frontend requires the most development moving forward—throughout this project we iterated quickly on design feedback without much regard to the quality of code. For example, on every “change in state” notification from the backend, all rooms are redrawn using JavaScript and JQuery. We plan to switch to a frontend framework such as React.js to better organize the code. Then, we plan to utilize the HTML5 <canvas> tag to create a scrollable world, in which any user only has a “viewport” into the space at any time. In general, we will need to clean up the UI before launching Talk Space as a marketable product.

4. Evaluating success

So far, our research has primarily been formative—the lessons from a particular design idea have led to new ideas. We made quick prototypes and mainly tested with small groups of friends to achieve quick feedback. In general, this subjective definition of success—“is this an experience that I imagined people would have in an online salon”—has helped us understand the essential elements of an online salon. Critically, we learned that rather than the structured nature of the sliders and fishbowls, the online salon requires the dynamic nature of Talk Space. I do recognize that if I were to continue to develop Talk Space into a mature technology, it may be useful to more methodically validate that Talk Space achieves the goals it set out to achieve. I would like to outline two approaches for possible future summative evaluation.

First, we could host and record a series of Talk Space interactions, and study them qualitatively. We could host discussions between a few people (fewer than ten), classroom or larger meeting size discussions (ten to thirty people), and much larger discussions (thirty to fifty people) to observe Talk Space in a variety of settings. We could observe whether behavioral cues have any effect; for example, look for evidence that boredom, or eavesdropping on a more exciting group leads to people joining a different group or creating their own. Within groups, we could look for standard social cues (i.e. body language) and whether they are used to negotiate who gets to talk during conversations. This would also be a good way to find unforeseen challenges in meeting the goals underlying Talk Space, or whether certain goals were incomplete or incorrect to begin with.

We could also design a survey for participants to better understand how they feel while interacting with Talk Space. Our overarching goal was to provide enough social translucence to help users find who they want to talk to. It is supposed to feel fluid, so that users can join, leave, and create groups easily. We could ask questions such as “were you able to find people with whom you wanted to converse?”, “how comfortable were you switching between groups?”, “was the ability to see and/or eavesdrop on other conversations useful?”, “what hindrances to conversations did you face?” The answers to these questions will help us gauge whether users “feel” how we would expect in an online salon.
5. Discussion

5.1 Higher layer protocols
Through Talk Space, we have explored the “qualities” of the online salon. However, Talk Space is meant to be a flexible platform for dynamic online conversations that can be extended to fit specific use cases such as online classes or office spaces. We expect participants to come up with implicit protocols for conduct (ones that evolve with no changes to the code), and we can also explicitly add features for specific scenarios.

My friends and I have already used Talk Space with implicit protocols for group projects. We noticed that we usually like to work alone, sometimes had to talk to one or a few other members of the team at a time, and occasionally meet with the entire group. An implicit protocol emerged that group members would create separate rooms in Talk Space, and stay in their own room if they did not need to talk to anybody else. However, if they had a question for someone, they would jump in that member’s room, and immediately be able to discuss. In general, we found that this led to fewer disruptions, and more productive conversations than Facebook chat (our usual form of communication). We imagine that more “implicit protocols” around how to behave will emerge within each community that uses a Talk Space, similar to what happens on communities on Reddit.

On the other hand, the virtual office platform Sococo could be imagined as a set of explicit protocols on top of Talk Space. Similar to our implicit protocols for group meetings, Sococo creates virtual office rooms for each member, as well as other spaces such as water cooler rooms and conference rooms. To enter a room, users simply click on its name, and can immediately start (audio) chatting. Office rooms can limit entrants, so a user may “close” their door to some or all other users. These rooms can be reimagined as Talk Space groups, but with specific names and access limits. Sococo calls being able to see all rooms and conversations establishing “presence”, while I have talked about the same concept in terms of social translucence.
We could use explicit protocols to support more structured discussions like the fishbowls. A challenge we noticed while designing the fishbowls was that no existing video chat platforms (Google Hangouts, Skype, etc.) allows users to fluidly switch between being in a discussion, and watching a discussion. Instead, we could utilize the dynamic groups of Talk Space, and build the queueing system discussed above. When the system decides to rotate somebody in and out, their video feed could automatically be added to or removed from the group.

5.2 It all comes down to the users
My intent is not to posit Talk Space as a perfect solution for online discussions. Though we have seen that Talk Space can be a valuable tool for online communication, like any other online platform, it can be abused. Chat Roulette, a site to talk to random people online, is constantly plagued with adult images that often push people away from the site. Conceivably, the same phenomenon could start occurring in Talk Space. At the same time, groups of people that really want to have productive discussions could use text-based approaches, or even our fishbowls or sliders. To a large extent, the success of Talk Space, or any medium of online communication, is dependent on how users choose to interact on it.

Rather than focus on how to police impolite behavior, we have focused on providing the mechanisms and social translucence necessary to hold a participant-structured online salon. We expect users to extend it with implicit if not explicit protocols. Our hope is that rather than Chat Roulette, Talk Space follows Reddit, in which users are free to create, subscribe, and unsubscribe from “subreddits”. Many subreddits have become essentially online communities with implicit social protocols, or explicit moderation. We expect that over time, communities will form within different Talk Spaces.
5.3 Technological Challenges
Throughout this project, we have assumed that the bandwidth and latency issues that underlie current video chat software do not exist, or will disappear in the near future. However, I discuss the issues here, and why I do not believe they will present a significant problem.

Traditionally, video chats use a client-server model. All clients send their video streams to a central server, and the server pushes all video streams to the clients. The server is the main bottleneck, and the clients’ distance to the server also dictates the quality of their video streams. However, Talk Space runs on a peer-to-peer model using WebRTC. In a nutshell, after logging into a central server, each client creates a separate audio/video connection to all other clients, forming a “mesh” network. While a mesh network cannot support many participants at the same time (we even saw lag with four participants over slow connections), it reduces the load on a central server. It also means that if participants are very close together (e.g. on the same LAN), or at least closer to each other than to a central server, then the network latency is lower, and video quality can be higher.

![A mesh network, where each client has a connection to all others](image)

We mainly tested Talk Space on the University of Washington campus with high speed connections (~50Mbps), so we faced no lag, even with ten (or presumably more) participants. We also noticed that in general, if even if participants were far away (e.g. across the country), as long as they had relatively high speed connections, we were able to chat without noticeable lag.

As more and more users join a Talk Space, clients would not have the bandwidth to create an audio/video connection to every other active user. Instead, since only a limited number of participants will be visible on a user’s screen at one time, users should only create connections to those visible in their current “viewport”. As users navigate through the space, they will only exchange data with the people they can see, and those that can see them—similar to how Google maps only downloads data relevant to the current screen as you move around or zoom into an area. Thus, we will efficiently use the available bandwidth without sacrificing the video quality.

5.4 Conclusion: Talk Space in context
This paper has envisioned the concept of an online salon—a space for quality, scaled discussions. Through various design phases, we learned the “qualities” of an online salon—rather than by crafting a strictly structured interface, online conversations are better scaled as in real life: by splitting into subgroups. The interface should be fluid enough to allow participants to move between, or even create new groups. It should preserve “social translucence” to help aid
these movements, as well as negotiate “who gets to talk” within one conversation. We finally created Talk Space, a flexible implementation of our online salon.

Web 1.0, which Tim Berners-Lee calls the “read-only” phase featured static pages to download and view, but there was little reciprocal interaction or contribution from a user. Then, Web 2.0 was then the “read-write” phase of the internet, in which users engaged with each other through mass text communication such as blogs, comments, and wikis—launching Facebook, YouTube, and Wikipedia.

However, as the Internet progresses forward, it is characterized by both increasing access to the web and much higher bandwidth. We are already witnessing Gigabit internet with initiatives such as Gigabit Seattle, and Google Fiber. At the same time, we are seeing rise in the use of video chats, and other content-rich sources such as gifs, videos, Second Life, and online gaming. The future of Internet communication is characterized by less client-server-client interaction, and more by smarter peer to peer interaction. Video chats between large numbers of people have never been possible in the past because of lower bandwidth—Google Hangouts and Skype still suffer from this problem. Talk Space takes advantage of the progression of Internet technologies, and is a step toward rich, scaled Internet communication.

6. Acknowledgements
I owe a huge thanks to Travis Kriplean, Michael Toomim, and Alan Borning for contributing their time and vast insights on this project. I would also like to thank the many friends who helped me test different prototypes and designs that were crucial in exploring the qualities of the salon.

Works Cited


