Social Virtual Reality Best Practices

Renee Gittins
July 30th, 2018
Version 1.2

Contents

Introduction

Virtual reality creates realistic, immersive experiences for its users. It also allows for freedom of interaction that is unprecedented in previous entertainment mediums. These features combine to create extremely powerful experiences that not only allow users to interact in a realistic manner with virtual reality applications, but with other users as well.

Because of the power of the virtual reality experience, developers should apply best practices to create the most enjoyable experiences possible for their users. Cases of harassment and abuse not only negatively impact the user, but can tarnish a project or developer's reputation as well. The tools in this document help reduce abuse and harassment, improving both user experience and public perception.

Moderation Layers

Moderation is a method for managing communities to ensure that behavior of the community members stays within defined bounds. Moderation can be handled in various layers. These layers can be combined to form more complete moderation coverage of content, while discouraging abuse opportunities or flaws that may be present in a single layer of moderation. Each layer and system needs to be evaluated in context of the content it will be applied to, as the moderation needs and potential moderation abuses of a competitive social VR game are very different from that of a casual VR social space.

Personal Moderation

The most accessible type of moderation for users is personal moderation. Personal moderation is the set of moderation tools and powers presented to users for them to curate their virtual experience. Due to the nature of these tools, personal moderation is the fastest acting layer of moderation, and the least subject to abuse and therefore requires the least amount of oversight to protect the community and integrity of the experience.

Standard Personal Moderation Tools

There are a set of standard personal moderation tools that are used widely in games and other forms of social online interactions. These tools include:

- Mute - Mute the targeted user's voice. This can be one directional or both directions, so that the muted user can no longer hear the voice of the user who initiated the mute.
- Kick - Remove the targeted user from the group or location that they are sharing with other users. This ability may be unique to the owner of the group (i.e. "party leader") or shared by everyone in the area.

- Block - Block all interactions from the targeted user, including messages, invitations, and party requests. Block can also be extended to matchmaking and even objects with which the blocked users interacts.

These standard tools can be expanded to include a concept that has an application more unique to virtual reality. The sense of presence and freedom of physical movements in virtual reality gives more weight to visual motions and actions of a user other games and online interactions. Thus, a similar personal moderation tool can be added for use in virtual reality:

- Hide - Make the targeted user invisible. This can be one directional or both directions (i.e. the hidden user can no longer see the user who hid them).

While the application of hide is much like that of mute, the type of interaction that caused one user to hide another (e.g. vulgar motions towards another player) is best reduced when the invisibility applied in both directions to protect the user and discourage harassment.

It is also important to consider additional interactions between players and tools that may avoid these moderation tools. For example, extra consideration would need to be taken if an in-game camera may capture the avatar of a user who should be hidden from the observer.

Personal Moderation Tools Abuse Considerations

These standard personal moderation tools are susceptible to various levels of abuse.

Muting cannot be abused when applied in a single direction, from user to target. There is a slight risk of both direction muting, as, in a group, it can enable the user to speak freely about the target to others in the group or area without the target's knowledge. Either implementation can cause frustration on the target's part, and may frustrate other players who are in the group, as core communication about the game will also be muted between two members of their team. Still, neither method should significantly affect either user's virtual reality experience. Muting is the lowest risk personal moderation tool when it comes to abuse while it still provides strong protection.

Kicking, especially when the ability to kick is shared by all members in a group or area, is more susceptible to abuse, particularly in competitive virtual reality games. If a game has groups that exist inside and outside of the competitive section of the game, it may be preferable to allow kicking only outside of the competitive section. For example, a user may be able to kick a person out of their party in a lobby area before the match starts, but not within the match itself.

Blocking is the personal moderation functionality that is most susceptible to abuse, but it also provides the greatest protection for users from abusers. Blocking can be abused in competitive games with random or ranked matchmaking. A user can block opponents they do not wish to compete against in hopes of gaining an advantage for achieving higher ranks. Furthermore, when a user blocks a significant amount of other players, they may end up in a situation where they are unable to find a match which does not have a player they have blocked in it already, resulting in infinite queue times. Thus, if blocking is implemented in a competitive virtual reality game with matchmaking, the development team should

consider allowing a blocked user to match and compete against the user who blocked them, while having other limitations automatically applied to them (e.g. mute) to prevent abuse or conflict.

Finally, when a third person is in the experience, the use of hide can allow one user to act inappropriately towards another they are hidden from without notice, while still being watched and even recorded by a third party. Hide also may give players an unfair advantage in competitive games and should be applied with consideration of potential abuse for the benefits of being unseen.

Enforced Personal Space

Enforced personal space is a powerful personal moderation tool that prevents other users from encroaching on the personal space of a user. The implementation of enforced personal space depends on the experience it is affecting. In general, enforced personal space prevents locomotion, such as teleportation, into that space by other users. It also provides a shield against other users and/or objects entering that space.

Enforced personal space can be implemented by fading-out a character when they enter the personal space of the user. This removes the ability for another player to be visually within the user's personal space and also prevents a the player from using their avatar to block the user's camera. It can also work in the both directions, fading-out the avatar of the user entering the enforced personal space defined by the user while simultaneously fading-out the enforced space user's avatar for the user entering the enforced personal space. This implementation helps users avoid unwanted interactions, both ones that the user and that other player can observe.

This fade-out technique can also be applied to physical objects that are thrown or held by other users. For example, in a billiards VR game, a developer may want to fade-out one player's cue stick when it enters the enforced personal space of another player to prevent it from being used as a distraction. Similarly, a developer may wish to fade-out the player whose enforced personal space is being entered by the cue stick, so that the aggressor cannot visualize the prodding.

An enforced personal space may also prevent other users from entering it completely, which provides a higher level of protection for users. However, this implementation can greatly affect gameplay and the range of user interactions. If it affects smooth locomotion such as walking, it may also cause situations that cause motion sickness.

Enforced personal space can be either a default size or with a size controlled by the user, including turning it off completely. The radius of the space can be controlled dynamically by the user or various pre-set size options can be presented to the user.

Enforced personal space cannot be applied to all social virtual reality experiences, as some experiences require close interactions between users. Close interaction is unavoidable in these experiences, though enforced personal spaced can also be applied to a small area around sensitive parts of the body, such as the groin and chest. Other methods of moderation may also be used to avoid or punish inappropriate behavior.

Interaction Level

Interaction level controlled moderation is a Personal Moderation system that can be adjusted quickly to fit the user's desired level of interaction at the time. This kind of system is particularly valuable for social applications that can be approached at a variety of intensity levels, so that the user can adjust their interaction level to match their current level of outgoingness and engagement.

The number of levels in the system and the exact permissions or restrictions on each level can vary, but this is an example of the states in a simple interaction level controlled moderation system:

1. Available - User is open to interaction from all other users, including messages, friend invitations, and party invitations.
2. Limited - User is only open to interactions from their friends. They will not receive friend invitations during this time, but can send them, and they will only receive messages sent by friends.
3. Unavailable - User does not want any interaction from any other user. They appear as "offline" in their friends' menus and will not receive any messages, invitations, or party requests, from friends or strangers.

Interaction levels are flexible and quick for the user to adjust. They best support social virtual reality experiences where varied levels of interaction are possible. They have reduced benefit for social virtual reality with a single level of interaction, such as ones that have only competitive matchmaking.

Community Moderation

Community moderation is management of the community and people within it by outlining and enforcing rules. Moderation of a community can be done actively through in-experience representation with active moderators or more passively through actions made in response to tickets. There are three systems that can provide community moderation: company hired moderators, outstanding members of the community who have been granted moderation powers, and a system that automatically reacts to complaints from the community, allowing for community self moderation.

Company Moderators

Moderators hired directly by the developer or publisher can be thorough and unbiased in their moderation of the community. Company hired moderators help control the tone of a virtual space and can be easily instructed to uphold specific values for the community. Company moderators are also particularly useful for reviewing items that might contain sensitive information about users, such as tickets and community analytics information, that would be inappropriate to share with members of the community.

Company moderators can act as representatives of the company and help encourage users to keep their behavior within the company's guidelines for interactions. Their inclusion during the launch of a product can help greatly to shape the initial community behaviors.

Company moderators can be expensive, especially for smaller studios, and can also make a community feel sterile or constrained. Community members will act differently with the knowledge that they are being watched, known as the observer effect or the Hawthorne effect, which will reduce abuse but may also affect their positive, casual actions as well. Negative effects of moderator presence can be reduced by ensuring that the paradigms of communication and cultural understanding of company moderators matches that of the community. Company moderators should be carefully selected as they will be seen as the face of the company within the experience.

Designated Community Moderators

Designated community moderators are members of the community that have been granted moderation powers. These moderators often do not require compensation, but have less accountability than paid moderators and therefore are more susceptible to abuse of their powers. Due to privacy regulations, these moderators should not be given access to any sensitive information about other members of the community.

Good community moderators will work tirelessly to foster a positive community and can be a huge boon to the community culture of the project, but unscrupulous ones can cause chaos and hurt the image of the project and company.

Community Self-moderation

The community can also regulate themselves through self-moderation. Self-moderation has to be watched for misuse or abuse, but it can greatly decrease the overall resource costs of moderation of particularly active communities.

Community self-moderation can be implemented in many different ways, three of the most common are:

- Allowing for removal of a user from an environment or party through "kicking". The ability to kick may be held by a single designated community member (e.g. the first person in the environment or the leader of the party) or it may be done on a vote-basis, where one player initiates a vote to kick another from the area and the other players vote in confirmation or disagreement of the removal.
- Banning, muting, or removing other functions from players who have met a threshold of abuse reports by other users in a certain time period.
- Allowing players to commend or report other users for their behavior and calculate each user's total behavior points, which can be used to lock or unlock functions for that user. These behavior points can also be affected by time or other variables to provide chances for redemption.

The greatest risk of community moderation is abuse by users. In competitive games, the best performing players are likely to be reported by their opponents without proper reason either maliciously to try to remove them from the competition or as an expression of frustration. Similarly, easily understood community self-moderation systems can be used by malicious groups of players to harass a targeted

user. If four reports in under 24 hours results in a ban, then four users can join together to ban anyone of their choosing.

Thus, when implementing a community self-moderation system, it is important to put in checks for frequency of reported compared to playtime and other checks to prevent abuse of the system.

Platform Moderation

Moderation can also be applied at the platform or store level. If a developer can provide sufficient evidence, or if the complaints against a particular user are consistent, the platform can institute a temporary or permanent ban. If a platform handles matchmaking of users, it can also apply reports about users to the handling of matchmaking similar to self-moderation systems.

Platforms can treat reports about community members similarly to how developers handle reports for community self-moderation, especially if they have a large community and want to implement a system that automatically takes action based on reports. The details the platforms require developers to provide in each report can further help refine their processing of and reaction to reports.

Environments

Social virtual reality can exist in various virtual environments. The handling and management of users across these environments can reduce and contain harassment. Environments can be seen as rooms in which users interact freely in virtual reality but could be implemented in other ways.

Reduced Risk Environments

A good way to ensure users have an enjoyable social experience is to place them in environments that are populated only by people they have marked as friendly. In fact, if players can only interact with those they have marked as friends or are friends of friends (e.g. being in a party with a friend as a leader and their friends as other party members), most moderation tools become unnecessary.

However, these reduced-risk environments are also very limited in their audience, preventing a vast majority of users from interacting with each other freely, due to the lack of connections. In a virtual reality experience with a small population of users, it is conceivable a user may never have a friend available to interact with in their environment. Additionally, this may encourage users to find connections they do not know through online forums or other places, which again raises the risk of abuse.

Designated Environments

Designated environments can give users freedom to interact with others within a range of explicitly allowed behavior for that environment.

Explicitly Designated Environments

Explicitly designated environments are environments which are labeled or defined as allowing certain behavior. For example, a common division of environments is "family friendly," which disallows swearing, rude gestures, etc, and "mature," which allows for these actions. These environments can be created and defined based on the needs of the community and the range of tolerance defined by the developers, similar to how traditional games will often have "player versus player" (or "PvP") zones to designate areas with a different set of interaction rules between players. Some developers may wish to avoid having a truly "no rules" set of environments, while others may want to embrace environments designated as such to allow users to act completely without restrictions.

User Defined Environments

Users can also play a role in defining the allowable behavior within an environment. Environments that are labeled by persistent names (e.g. "Bravo Room") or even just numbers, can develop their own culture and levels of tolerance for certain behaviors, even if it is not explicitly defined or listed. Giving users an option to select the environment they wish to enter and to change environments after entering one, will reduce the number of abuse incidents slightly because the user can exit or alter the discomforting environment.

This can be further accelerated and enforced by using designated community moderators who control the environment, setting the rules they determine the best fitting and upholding them through use of their moderation abilities. This is not a perfect solution, as it is still susceptible to abuse and lack of constant moderation or conflict between different moderators attempting to moderate the same environment.

Giving a single user control of the rules of the environment also can help support a range of environment behaviors and reduce friction between users. Either users can generate and list their own environments with specific rules, much like private servers for traditional online games, or simply use moderation powers to enforce rules they wish to uphold. The first user in an environment can be given moderation powers over that environment and uphold their own set of rules.

Exile Environments

When using a behavior-point based community moderation system or any moderation system that tracks numbers of complaints against a user, it is possible to greatly reduce the amount of abuse the average user encounters by implementing exile environments. Exile environments are environments that are designated for users who have shown or been reported for bad behavior. These exile environments allow for users who have been reported to be abusive users to continue participating in

the virtual reality experience while only allowing them to interact with users who have been reported for similar behavior. This exiles toxic players to an area where they can freely interact with each other while not upsetting the rest of the user base.

Culture Development

Toxicity and abuse in a virtual community can be a core behavior of the community's culture that cannot be easily changed even with the best technical approaches and moderation tools. Fostering the desired culture of a community is key to creating an experience that users can enjoy without unwanted harassment. It also greatly affects the first impression an experience makes on both users and the media.

Code of Conduct

A Code of Conduct is a low cost method of informing a community of the allowed behavior. A Code of Conduct is generally a list of behaviors that are considered acceptable or unacceptable. Not only does it inform users of what actions are appropriate within the experience's community, but it also empowers community members to stand up against inappropriate behavior by giving them explicit guidelines to point to when someone acts outside of them. It can also be held as reason to remove a player from the community in the End-User License Agreement (EULA), though often the ability to remove any player without reason is included in EULAs to avoid legal disputes over the matter.

A Code of Conduct should be short, easily accessible, and written in clear language to be most easily consumed, understood, and upheld by users.

Content Consideration

The content of a virtual reality experience can have a great affect on its culture. If imagery of the user avatars accentuates their sexual aspects, it is likely that there will be more sexual comments and interactions, and likely more sexual harassment.

Imagery, language, symbols, jokes, and events in a virtual experience affect how users perceive the experience and define its culture, including what is considered appropriate. Identifying and removing imagery and language that make suggestions against the desired culture of the experience can help set the culture on the desired track from the start.

Promotion and Celebration of Behavior

Another way to assist with cultural development of a virtual community is to public promote and celebrate positive behavior. This can be done in a wide range of approaches, from congratulating community members on social media to granting profile badges commending upstanding behavior. Even a simple notification once a week informing users who have not been reported that week that they are

appreciated as upstanding members of the community can act as positive reinforcement to continue their positive behavior.