Technical Disclosure Commons

Defensive Publications Series

January 2022

Enhanced Telepresence via Sensory Augmentation

Wim Deweerd

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Deweerd, Wim, "Enhanced Telepresence via Sensory Augmentation", Technical Disclosure Commons, (January 11, 2022)
https://www.tdcommons.org/dpubs_series/4834

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Enhanced Telepresence via Sensory Augmentation

ABSTRACT

Telepresence interactions can appear unnatural and alienating, thus making them feel superficial, hampering the development of meaningful human connection between participants. Current mechanisms to augment telepresence are inadequate for overcoming these sensory deficiencies since these typically augment only a single sense and sometimes result in attenuating essential sensory experience and/or exaggerating the sensory impact. This disclosure describes augmentation techniques to create a more natural sense of presence during video conferencing by adding various layers of natural background visuals and/or sounds, with specific user permission. An indication that a particular participant's appearance has been augmented is provided to all participants in the video conference. A suitably trained machine learning model can be used to select the types and intensities of background audiovisual stimuli to which a user is the most responsive.

KEYWORDS

- Video conferencing
- Autonomous Sensory Meridian Response (ASMR)
- Sensory augmentation
- Background sound
- Background visual
- Visual appearance
- Enhanced appearance
- Ambient noise
- Telepresence
- Augmented background
- Image filter

BACKGROUND

People increasingly rely on video conferencing for professional and personal communication. During a video conference, the interactions can appear unnatural and alienating, making the users think of the interaction as superficial and thereby hampering the development of meaningful human connection between people in the video conference. To overcome this limitation, some video conferencing applications provide users with the ability to augment their telepresence via various mechanisms such as setting the background to a chosen image, reducing unwanted background noise by using noise cancelation, applying filters to alter one's appearance, etc.

Such mechanisms to augment telepresence typically focus on augmenting only a single sense, e.g., visual, auditory, etc. For instance, adjusting the background affects only the visual appearance. Moreover, the augmentation can sometimes result in attenuating essential sensory experience. For instance, noise cancelation can increase the sense of isolation because of reduction in ambient environmental sounds. In contrast, mechanisms such as filters can exaggerate the sensory impact and are often used for purposes other than enhancing the naturalness of telepresence interaction. Because of these factors, current mechanisms for telepresence augmentation are inadequate to overcome the sensory deficiencies in video conferencing. In fact, current mechanisms can make the unnaturalness of the interaction more, rather than less, apparent.

For many people, stimulating one sense can result in experiencing sensory perceptions related to a separate non-stimulated sense. The phenomenon is termed as Autonomous Sensory Meridian Response (ASMR), also known as synesthesia or frisson. ASMR effects can occur in different forms of sensory combinations, with each experienced by various percentages of people

in the general population. For instance, ASMR responses can be induced via occasional sounds and/or visual stimuli.

DESCRIPTION

This disclosure describes techniques to enhance the naturalness of telepresence during video conferencing interaction. If users permit, their telepresence during video conferences is augmented by artificially adding various layers of natural background visuals and/or sounds to create a more natural sense of presence. Such augmented telepresence can induce ASMR that can improve the connection of contentedness among participants of a video conference. An indication that a particular participant's appearance has been augmented is provided to all participants in the video conference.

Inducing appropriate ASMR can be achieved by augmenting telepresence with the insertion of one or more background visuals and/or sounds such as:

- Quiet, repetitive sounds resulting from someone engaging in a mundane task (e.g., turning the pages of a book, sharpening a pencil, tapping on surfaces, etc.);
- Someone in the background attentively executing a routine task (e.g., preparing food, making coffee, etc.);
- Common sounds generated by various actions of persons (e.g., chewing, crunching, slurping, biting, pouring liquid, etc.);
- Receiving personal attention (e.g., an artificially added character appearing to be looking into the camera);
- Initiating the stimulus through conscious manipulation without the need for external video or audio triggers;

- Hand movements, especially onto the face, by an artificially added character appearing in the background;
- Various types of music;
- Muted noises and soft sounds from the street, stores, etc. (e.g., boxes being stacked, personnel chimes, vehicle traffic, etc.);
- Rhythmic, dynamic, harmonic, and/or melodic violations of a person's explicit or implicit expectations associated with musical frisson as a prerequisite;
- A whispering voice after an action, such as an e-commerce transaction, to provide positive reinforcement of the decision (e.g., to avoid buyer's remorse after a purchase); etc.

Such effects are added with user permission as a layer to a user's telepresence feed within a video conferencing interaction. Others interacting with the user via the medium receive the user's feed in which the background contains the inserted layer of visual and/or sound stimuli, thus potentially leading to perception of the interaction as genuine, realistic, and positive. Such perception can in turn increase engagement in the interaction and facilitate a greater sense of human connection.

If users permit, individual responses to various stimuli inserted in the background layer can be captured over time and analyzed using a suitably trained machine learning model. The output of the model can indicate the types and intensities (e.g., volume of sounds) of background visuals and/or sounds to which a person is the most responsive and favor those stimuli in future interactions. For instance, if the model output indicates that a person responds well to the background visuals and sounds of a café, stimuli that simulate a cafe are more likely to be used in future interactions at the person's preferred intensity levels.

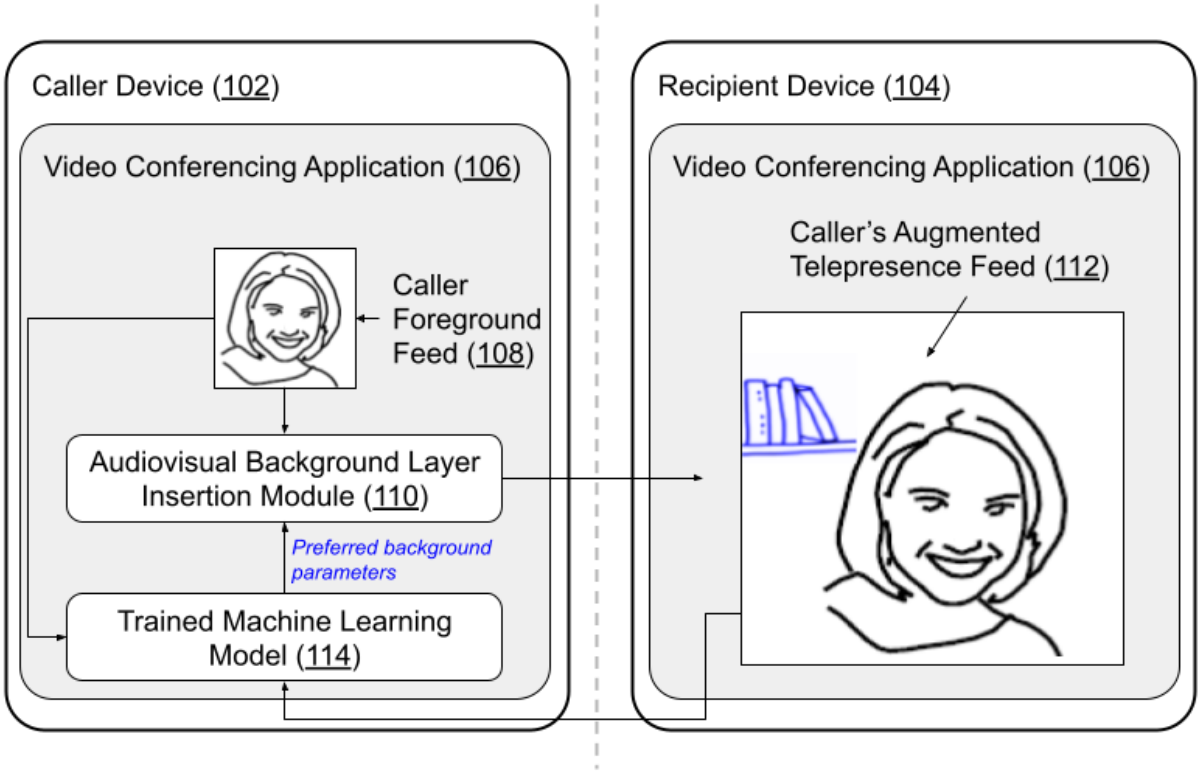

Fig. 1: Inserting visuals and/or sounds as background layer to augment telepresence

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. A caller uses device (102) to interact with another person via a video conferencing application (106). With permission, the video feed of the caller's foreground (108), shown in the left-side portion of Fig. 1, is augmented with the insertion of appropriate visuals and/or sounds as a background layer (110). The recipient device (104) receives the caller's augmented telepresence feed (112). For instance, the right-side portion of Fig. 1 illustrates that the recipient sees the caller against the background of a library. Sensory responses of the users to the augmented background layer are input with permission to a trained machine learning model (114) to determine the parameters (e.g., type, intensity, etc.) for the background that users prefer. These parameters can be used to adjust the background layer for future interaction.

If users permit, the techniques described above can be implemented within any application, platform, or service that provides video conferencing functionality with the capability for augmenting background visuals and/or sound in the audiovisual feed or can be implemented as a separate augmentation plugin. The techniques can support video conferencing calls for any purposes, such as personal communication, professional interaction, e-commerce transactions, etc. For instance, the addition of ASMR responses in e-commerce applications can help create more engaging and meaningful customer experiences that go beyond the typical limited interaction when shopping online via a web browser or an e-commerce app. Implementation of the techniques can stimulate ASMR in a non-intrusive manner, thus enhancing telepresence to create a more natural and realistic connection and having a positive impact on the overall user experience (UX) of video conferencing.

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's audio and video feed, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes augmentation techniques to create a more natural sense of presence during video conferencing by adding various layers of natural background visuals and/or sounds, with specific user permission. An indication that a particular participant's appearance has been augmented is provided to all participants in the video conference. A suitably trained machine learning model can be used to select the types and intensities of background audiovisual stimuli to which a user is the most responsive.

REFERENCES

1. Valliath, George T., Eric R. Buhrke, and Mark A. Tarlton. "Method system and apparatus for telepresence communications utilizing video avatars." U.S. Patent 8,072,479, issued December 6, 2011.
2. Saraiji, MHD Yamen, Shota Sugimoto, Charith Lasantha Fernando, Kouta Minamizawa, and Susumu Tachi. "Layered telepresence: simultaneous multi presence experience using eye gaze based perceptual awareness blending." In ACM SIGGRAPH 2016 Posters, pp. 1-2. 2016.
3. Vincent, James. "Zoom Escaper lets you sabotage your own meetings with audio problems, crying babies, and more." https://www.theverge.com/2021/3/15/22331744/zoom-escaper-sabotage-meetings-fake-audio-problems accessed October 11, 2021.