TELEPRESENCE ROOM ACOUSTICS

Ø Bredvei TANDBERG, Lysaker, Norway
TV Tronstad
JL Nielsen

1 INTRODUCTION

The term "telepresence" has been around for decades, and has been used for various ideas and applications. In the video conferencing industry this term has been quite recently adopted. It is now a buzz-word widely used for the top range systems offering the best quality of experience.

Telepresence conferencing has been defined as[1]: "the science and art of creating visual conferencing environments that address the human factors of the participants and duplicate, as closely as possible, an in-person experience". The human factors include all aspects of the experience, including but not limited to video and audio.

Development in video conferencing and network technology in recent years has made telepresence conferencing a practical reality, and the first generation of systems has been on the market for some time. There is a great expectation in the industry that telepresence developments will improve end-user acceptance and increase usage. It has also lead to new players emerging with products on the market, most notably large network equipment providers.

The typical application is shown in Figure 1, where 4-6 people sit on one side of a table and the conferencing system is giving an illusion of meeting other people around the same table. Electro-acoustic components are microphones on or close to the table and loudspeakers below screens.

This paper will list some challenges in telepresence system design, report a few lessons learned from early adopters, and briefly discuss room acoustics compromises.

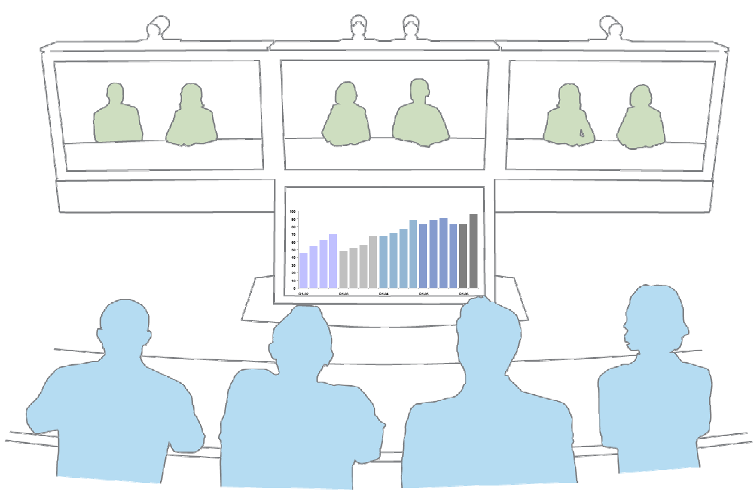

Figure 1: Typical telepresence application

2 TELEPRESENCE SYSTEM AUDIO

The design of the microphone system in a telepresence setting is initially challenging. The sound quality should be as natural as possible. Comb filter effects from nearby surfaces and shadowing from laptop PCs should be avoided. There should be no audible artefacts like noise, distortion, echoes, gating from mixing or signal routing algorithms, or RF interference from cell phones and other wireless devices. The design is further complicated by the strict requirements for visual appearance, and that the technology should be hidden, especially from the camera views.

Two-way loudspeaking real-time communication with the added delays of codecs and IP networks puts high requirements on the echo cancellation technology employed. State-of-the-art echo control algorithms have to be used. The main feature distinguishing telepresence from mainstream video conferencing is a requirement of multi-channel audio providing a sense of direction. This further complicates the echo control. Making multi-channel echo-cancellation work in practice without artefacts is a major challenge[2].

Therefore, and due to the extended geometry of the situation and still limited bandwidth and processing resources, the directional audio algorithms employed in most first generation systems are based on quite simple and straightforward panning. 3D sound technology, which might easily be connected with the term telepresence, is not used. However, the panning algorithms are more than adequate to provide the necessary directional information, especially when supported by high quality video.

Working compromises for the microphone system requirements can luckily be found. Telepresence microphones can be chosen directional and mounted in fixed positions fairly close to the meeting participants, thus maximizing sound pickup quality while minimizing acoustic feedback. This limits the sound pickup range, but is acceptable due to the fact that the meeting participants have to sit in dedicated positions to appear correctly on camera.

On the reproduction side the main challenge is to find a good compromise between loudspeaker placement, video requirements and visual design. Screen and camera placement is important to optimize image size and the sense of eye contact between meeting participants. This invariably leads to loudspeakers being placed quite low, which may compromise the directional audio performance. However, this has been demonstrated to function reasonably well in the full experience with multiple modalities supporting each other.

3 ROOM ACOUSTICS DESIGN

Rooms for telepresence conferencing are often used for long meetings, hence the environment is very important. This includes of course the acoustics, but also lighting, air quality, temperature, etc. The basic room acoustic requirement is that it should be a pleasant environment providing easy communication with no audible artefacts.

3.1 Room reverberation

The first generation of telepresence systems has been accompanied by recommendations for rooms with lots of sound absorption. This thinking goes along the lines of recording or broadcasting studios or listening rooms, focusing on maximizing the auditive experience of the transmission system. This will ensure good quality and clarity on the sound picked up and reproduced.

However, with this choice other aspects of the experience are overlooked. Such rooms can easily end up having a reverberation time around 0.2-0.3s, and it becomes a tiring task to have natural conversations for extended periods of time. In addition, the seating plan along one side of the long table reduces the visual and acoustic contact internally in the room.

The challenge is therefore to design a room that works well for both situations:

1. Communication internally in the room
2. Communication between different rooms linked by the telepresence system

Testing workshops with developers and feedback from some of the first generation system customers show that 1. should have at least equal weight as 2.

Weighting internal conditions heavily by increasing acoustics support without compromising speech intelligibility leads to a suggested reverberation time range of 0.4-0.8s with diffuse conditions. However, a compromise between 1. and 2. must be sought, leading to a reverberation time recommendation of 0.4-0.5s.

In addition we suggest that excess low-frequency reverberation or strong modes should be avoided, and that the energy of reflected sound actually rises slightly around the higher frequencies important for speech intelligibility.

Attaining this RT should not be a problem in normal design of a meeting room of around 100m^3, which can be a typical size for a telepresence room. However, systems will have to be installed in all kinds of rooms, also rooms where far from all surfaces may be available for acoustic treatment. Also, such treatment will have to be made adhering to strict visual design requirements.

The back wall is probably the most important surface to get sound absorbing and/or diffusing, but this is very difficult as it is the primary surface filmed by the HD cameras. It's light and texture properties will always take priority. The side walls are not visible in camera views and is in principle available for acoustic control.

However, ceilings are often highly absorptive, and they are the surfaces that customers often cannot or will not change during installation. This invariably leads to reflective side walls to reach the suggested reverberation time. If the customer has a reflective ceiling that cannot be changed, an area corresponding to both the side walls would have to be highly absorptive.

Diffusing elements can be very hard to incorporate due to visual design, and avoidance of large parallel surfaces can be impossible due to lack of space in the room.

The combined consequences of these challenges may be that the necessary room reverberation can be achieved, but at the cost of not being able to diffuse the sound field properly.

3.2 Early reflections, echoes and artefacts

To make a pleasant sound environment, strong reflections creating comb-filter effects, directional artefacts, flutter or echoes should be avoided. However, with the frequent restrictions on the walls this may be hard to avoid. In addition the telepresence system itself can be a major plane and reflective surface creating special problems. Together with the typically non-absorptive back wall it creates an excellent source of rapid flutter.

As the table and seating plan is wide and one-sided, communication between opposite sites should be aided by sound reflections. If the ceiling is not available for tailored reflections, the side walls should be kept somewhat reflective. This also increases risk of flutter effects.

Whether the first side reflections from the loudspeakers to the participants should be attenuated by absorbers or not is also an open question. To minimize the risk of upsetting the directional sound, that part of the side walls should be somewhat absorptive. However, in the round-table setting the system is set up to simulate the reflection is actually present.

3.3 Background noise

Part of creating a good sound environment both internally and between linked rooms is to strive for a low noise level. The telepresence experience is also dependent on well controlled lighting, therefore windows should not be present in the room. As the system screens and equipment, the lighting and the persons in the room will generate quite a lot of heat, the room typically ends up being equipped with an air conditioning system, or at least connected to a high capacity ventilation system. These should be designed with care to avoid increased background noise.

Sound insulation can be important as well. Not only are the rooms often used for classified meetings, but external noise may be more detrimental to microphone sound quality as such noise often is non-stationary and therefore difficult to attenuate electronically.

4 CONCLUSIONS

After the first generation of telepresence systems have been in the market for a while, it is clear that room acoustics design for telepresence is more challenging than the first recommendations show. Care must be taken to make communication in the room natural and effortless during long meetings, while keeping the pick-up and reproduction of sound for communication to the far-end as high quality as possible. This in itself is a challenging compromise. It is further complicated by stringent visual design requirements, and the often overlooked fact that ceilings are in general not available for acoustic optimization.

A compromise seems to be to make the room somewhat more reverberant by having mostly reflective walls. This increases the support for communication internally in the room, at the cost of some acoustic artefacts. More experience with such rooms is necessary to conclude on the optimum compromise.

Further development in absorber and diffusor technology would aid optimization of video, audio and visual appearance, and would be very welcome.

Human factors most emphasized in telepresence design tend to be those most important for first impressions like visual design and video quality, which may be necessary for getting the systems sold and deployed. Factors like audio and acoustics are and must be high quality at present, but we see potential for further improvement. It is our expectation that such advances in telepresence sound quality will make the long-term aspect of the telepresence experience more natural and the effort of long meetings less, make users more satisfied and customers come back for more.

5 ACKNOWLEDGEMENTS

We thank Magne Skålevik from Brekke & Strand Acoustics for fruitful discussions on room acoustic requirements for telepresence.

6 REFERENCES

1. H.S. Lichtman, “Telepresence, Effective Visual Collaboration and the Future of Global Business at the Speed of Light”, Human Productivity Lab, Aug. 2006.
2. J. Benesty, D.R. Morgan and M.M. Sondhi, ”A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation.” Proc. ICASSP97, pp. 303-306 (1997).