COMBINING NATURAL ACOUSTICS AND AUDIO IN A DEBATING CHAMBER

R Essert   Imagine Sound, London, UK

1   INTRODUCTION

We consider the complex acoustical situation in a debating chamber where speakers and listeners are spread around the room and there is a distributed sound reinforcement system.   This paper is a briefing document, with observations and lessons learned, concepts and priorities.

This work stems from our work with the Canadian Parliament in Ottawa, but the issues it faces are common in whole or in part to other debating chambers. There are several hundred seats in an oppositional seating arrangement, and a protocol that every MP speaks to the Speaker of the House means the talker's voice is aimed toward the few and away from many. In Ottawa, as in Westminster, the official protocol is that the chair recognises one person at a time to speak. Much of the time there are just a few Members in the chamber reading statements into the record.   But when the chamber is full there are times where the responses mask the speeches.   So, in some cases the interference from others will impact the speech intelligibility in the room and in the feed to others outside the room.

How did MPs project before there was sound reinforcement?   Like earlier actors they spoke more slowly and strongly. They learned to be heard and understood through experience, and architecture and democracy evolved. Today parliamentarians want to take advantage of audio support, to reach more people and to place less strain on their voice.   Some assembly halls operating today were designed before sound systems, and the architecture has remained more or less as the original, while audio systems have been introduced. Even if acousticians could devise means to improve the natural acoustics, protected architecture and materials in historic buildings can limit the possible improvement.
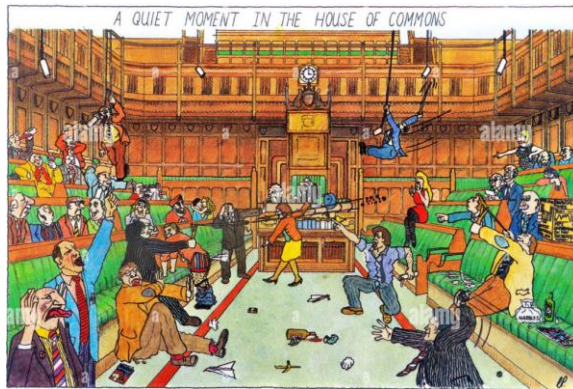
"When Parliament was first broadcast, for the first three days the BBC broadcast everything that came through the loudspeakers. It was libellous, it was unbelievably crude, but it was hilarious. The BBC panicked and said, "Somebody will sue us for libel. If it is in Hansard it is okay, but if it is not in Hansard we will be done for libel." So the BBC stopped broadcasting everything; now, it jams the broadcast so all people hear is, "Hear, hear, hear." It is terrified of being sued for libel.

The Chamber sounds like animals in a zoo, but for the people in the arena who can hear, it is often witty and sometimes caustic and destructive of careers—but that is politics. It is a rough old trade. We have to find some way of getting that across so that the public can get a taste of what is happening without it denigrating Parliament." – Mr Joe Ashton, MP   Hansard, 4 June 1997.

2 COMPLEXITY OF THE ACOUSTIC AND ELECTROACOUSTIC SCENE

2.1 Overview

Often the room acoustics and the audio processing are addressed separately, and by separate designers. In this situation the two must be considered in an integrated way. The large numbers of talkers and listeners spread through the room is a major complication. But considering first any one source (talker) and listener combination there are multiple acoustic and electroacoustic paths, the principal ones as follows:

1.   Natural acoustic -- direct and reflected /reverberant sound
2.   Sound transmitted a short distance to one (or more) microphones -- direct to mic, reflections between talker and mic. This includes mic directivity and placement, nearby finishes, talker movement
3.   Sound picked up from the microphones and emitted through multiple loudspeakers -- direct and reflected/reverberant contributions.
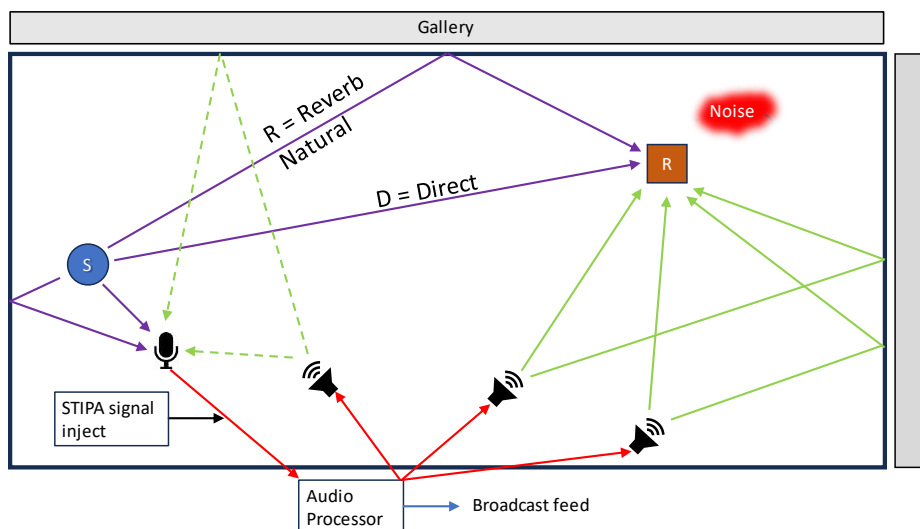4.   Feedback loop from loudspeakers back to open mic -- direct and reflected.

Figure 1 Multipath electroacoustic situation

2.2 Natural Acoustic

The natural acoustics of the room of course depend on scale, shape, materials, occupancy and locations of sources and receivers – here talkers and listeners. In a room with a reasonably linear decay EDT is an accepted metric for decay from a single natural source. That applies to one talker at a time. When there are multiple sources in a distributed system, EDT is not so helpful, except as a gauge that you might have trouble.

The talker may be anywhere in the room, and the intent is communication to everyone else in the room, plus the public gallery. The strength of the talker is relevant. Depending on the size of the space, some strong voices may be able to make themselves heard in the natural acoustic, but it is essential that all, even those with weaker voices, are heard. Directivity of the voice is just as important; the intelligibility behind a talker is far less than in front. In most large assembly spaces the protocol is for the person speaking to stand if they are able, and many rotate as they speak to

direct their voice to more of their audience – a natural reaction to the room layout.   This may help the natural acoustic, but may challenge the microphone pickup.

2.3   Distributed Audio Reinforcement

In a large assembly hall such as a national parliament there may not be sufficient strength for purely natural acoustic communication.   Parliamentarians in Westminster are used to being packed in a small chamber, designed at a time before audio assistance, and they still wish they could manage without it.   They need a bit of lift for loudness, but want the apparent source to be at the current talker, not at a central cluster. This has led to distributed loudspeaker and microphone systems. In Ottawa the loudspeakers are in the desk in front of each MP, facing away from the centre line, and in Westminster the loudspeakers are shared between two in the benches facing forward toward the centre line. In each case they are less than one meter from the ears of each listener.

Generally, people in an assembly or debate want to be in the same space, sharing the same acoustic with their opponents and colleagues.   The discussion is fluid, body language is involved, and multiple voices are simultaneously audible without crude software gating . (We have all learned over the past three years that those attending in person have an advantage over those who attend virtually. Debate is competitive, and it is theatre.) Achieving natural balance and imaging between the room acoustic and the electroacoustics is important to the sense of being in the event together.

With a distributed loudspeaker system in a reverberant space, as in a church, the energy put into the room builds and is spread in space and time. The direct sound and early reflections may not be loud, but the reverberant field can be too strong. The room acoustic needs to be dry enough that the many loudspeakers don't excite the reverberant field too much when they are loud enough to give local support. There needs to be a sophisticated zoned switching and muting system to control feedback and for each loudspeaker to deliver strength and delay appropriate for each zone around the room without exciting the reverberant field any more than is necessary. In Ottawa the gains and delays from current talker to the multiple zones of listeners are keyed to the active source location.

It is a given that the number of open mic channels must be limited.   So matrix switching systems have been developed that allow an audio operator to open one mic at a time when each talker is recognised. In Ottawa the microphones are in each desk and in Westminster they are suspended overhead, with somewhat more coarse coverage.   Pickup of speech from distributed microphones can be switched to select the person talking. In parliaments this is typically done manually by an operator, and in smaller spaces it might be automated.

2.4   Audio Broadcast Feeds

The quality of broadcast audio output must be high to meet international standards and audience expectations.
Tonal quality and pickup pattern are important, and can be customised. The mic switching for in-house reinforcement directs the stream feeds as well. But mixing and signal processing needed to achieve excellence will be different than what is needed for the in-room sound.   Broadcast streams are not mixed to binaural or surround formats yet, but this will be a good application when the technology and market have caught up. We shall see whether the directors of the stream find such opportunity attractive or beneficial.

The feeds for official transcription such as Hansard (by human and/or machine) can be inside or outside the room. In Ottawa the Hansard transcribers are in the centre of the chamber and have their own local loudspeakers. They want to be immersed in the action.   In multilingual proceedings, as in Ottawa, feeds to interpreters are also derived from the output.

The speech stream needs to be very clear to minimise errors in transcription and interpretation, and to give the best rendition to the audience at home.   Anomalies in the room acoustic can affect the

sound picked up by the microphone.   While only one person should speak at a time, that person can be anywhere. Consistency of physical distance and direction between mic and talker is helpful to achieve consistency of audio quality from one talker to the next, but it is hard to design this into the system. This is more audible for the stream than in the room.

Bleed between mics can muddy the broadcast feed. Sometimes 2 mics are open, eg Speaker and Prime Minister.

Spatial audio is not traditionally used for these applications, but it would seem helpful for the interpreters to have spatially distinct sources. There is technology available that can locate close-miked sources in a virtual spatial sound field.   Managing the virtual location for the microphone may involve choices relating to the video viewpoint.

3   SPEECH TRANSMISSION INDEX (STI) AS A QUALITY METRIC

As a metric STI does not seem to address this complex multi-modal sound field. One can measure the STI of a speech received through a complex system, but it is not something that can realistically be modelled. One could, in principle, develop a multi-level model comprising modules for architectural acoustics and electroacoustics patched together in a DSP framework such as Puredata or MaxMSP.

STI is a single channel measurement. It can be measured with an omnidirectional loudspeaker or with a voice source speaker with directivity similar to a human voice. It is important that the most relevant, not the most expeditious, choice is made in modelling and measurement.

The current measurement standard is IEC 60286-16 (Edition 5). Since the 1970's when STI was proposed by Houtgast and Steeneken, there has been quite a lot of work on intelligibility in binaural systems, and on output from modelling systems, and these should be connected to improve real-world prediction of intelligibility.   Various researchers, including Bronkhorst & Plomp have shown that binaural listening with source separation provides improvement in speech intelligibility in the context of interfering speech of 3 to 8dB [1,2]. Should this not be accounted for in our situation? In a sound field this critical and this complex it would be valuable to design and assess with spatial processing and rendering.

STI has come under scrutiny from several directions over the last 50 years, and has benefitted from improvements:
• Level dependent masking – the documented decrease in intelligibility above 80dBA has been included in later versions of the STI Standard.
• Speech spectra and forward masking functions of the ear and brain.[3]
• The frequency weightings of the different bands have been adjusted better to align with real voice spectra, and the female test spectrum has been omitted, leaving the worst case male spectrum
• Recognition of the effects of compression in hearing aids and stream compression
• Recognition that STI is not sufficiently degraded by discrete echoes.[4]
• STIPA for public address systems was included in 2003. RASTI has been declared obsolete.
• The concept of speech shaped noise has been added

The adjustments seem to make it more relevant for design of PA systems with close mics and few loudspeakers, but not much so for the debating chamber situation.

4   REAL WORLD COMPLICATIONS

Background noise from mechanical and electrical services can be relevant for natural acoustics, but for many situations the interfering noise is not continuous, but is the dynamic cacophony of others in the room speaking. The dynamics of interference are not accounted for in STI.

Most of the background noise is other voices. The spectral overlap between target voice and interference is known as energetic masking – intelligibility is reduced as part of the spectrum is masked. This is nominally accounted for in the STI method.   However, STI doesn't account for "informational masking" that occurs when the noise is interference from different speech. This arises from cognitive delays or errors in separating the newly arriving sound from the surrounding interference. [5]

There are additional complications for participants listening in a second language. If the sound of the talker is amplified throughout the room, those people listening to simultaneous interpretation have to listen through earphones, while the natural sound and PA sound become interference.   (The Canadian Parliament is bilingual in English and French; the language changes from person to person, and even within one speech.)

The challenge is made even more complex with the priorities of audio (and video) streaming in and out of the live debate.   The mass leap into virtual meetings in 2020 had huge consequences in convening parliaments. Patched-together AV systems were fraught with problems and limited the efficiency of governing. Linking people through videoconference still needs serious work, not least on the audio side.

We need a quality descriptor for the clarity of transmission from talker through to broadcast output that takes into account more than just in-room STI or electronic channel performance. Achieving good acoustical quality in the room does not necessarily deliver good broadcast audio, and sometimes broadcasting quality is the more precious of the two.

Design of the room includes some opportunities to bring together the acoustics and the audio system design, but in a heritage building there can be serious constraints on both geometry and finishes. This puts more of an onus on the audio system and its control. The starting point must be the in-person relationships of people to each other.

5   FUTURE: IMMERSIVE AND EXTENDED CHAMBER

Watching the UK and Canadian Parliaments on TV during COVID, and how they struggled with AV technology and virtual attendance, we began to appreciate the role of the architecture, the layout of the people in the chamber and the acoustics and audio on the quality and efficiency of governance.

We have a vision where the virtual participants, from home or other safe place, feel part of the group, where the sense of presence conveyed by the visual and aural spatial relationships between occupants and space are helpful to the quality and efficiency of the debate.

Consider a virtual parliament where the best aspects of the sound in the real House are retained in the virtual acoustics, and improvements could be made where appropriate.   Or perhaps an absence of chamber acoustic would be preferred. Either or both is possible. In any case a distributed apparent spatial arrangement of participants in virtual space would be an immense improvement over Teams or Zoom with their monophonic audio and checkerboard video arranged by some unknowing algorithm. When the spatial aspects of audio systems are included the different sound sources (real or virtual) can be spread all around the listener, so each speaker seems to be located in a distinct location in the listener's head space. With such spatial distinction the words are clearer and meaning registered easier in the mind of the listener.

For each listener, separating the information from the source you want to hear from other sound in the space involves quite complex processing in the brain. If a talker and the competing sounds are in the same relative direction, the brain has to rely on the differences in timbre or tone (voice recognition). This occupies much conscious attention -- and therefore brain processing -- just to hear, leaving little power for listening and thinking. Apparent spatial separation of sources from the listener's point of view helps to optimise intelligibility because our brains can spend less effort to understand who is talking and what they are saying, leaving more brain power to listen and think. Inside the Commons chamber Members benefit from this to some extent, as they can see and hear each other, assisted by the tight layout and, as noted above, the compact dimensions. The virtual connection could facilitate a greater sense of connection between the present and virtual representatives and their public.

6 CONCLUSION

The design and tuning of the audio system are key elements in achieving:
- sufficient speech intelligibility for those in the room for participants and audience
- intelligibility for real-time transcription and simultaneous interpretation
- subjectively appropriate balance between natural and electroacoustic lift in the room and stream sound that "matches" the video.

Predicting and measuring speech intelligibility in this situation involves a number of variables:
- multipath model with acoustical and electroacoustic paths
- considers the sensitivity of speech intelligibility to the various components.

A client brief for a debating chamber should not be limited to T30, SPL, STI and background noise criteria, but should embrace the interaction and complexity of the natural acoustics, the natural sounding reinforcement and the participation of virtual attendees. Spatial audio should play a part in the streaming of parliamentary proceedings.

When a building for debate among hundreds of people can be developed with acoustics and electroacoustics at the heart of the design, there is a chance to provide good clarity connection among the participants and the audiences.

7   REFERENCES

1.   A. W. Bronkhorst and R. Plomp, The Effect of head-induced interaural time and level differences on speech intelligibility in noise, J. Acoust. Soc. Am. 83(4) 1508-16 (Apr 1988).
2.   A. W. Bronkhorst and R.Plomp, Effect of multiple speechlike maskers on binaural speech recognition in normal and impaired hearing, J. Acoust. Soc. Am. 92(6), 3132-39 (Dec 1992).
3.   G. Leembruggen, M. Hippler, P. Mapp, Exploring ways to improve STI's recognition of the effects of poor spectral balance on subjective intelligibility, Proc. IOA, Vol. 31 Pt 4, 133-169 (2009).
4.   R. Hammond, P. Mapp, A, Hill, Disagreement between STI and STIPA measurements due to high-level, discrete reflections, Proc. AES 142nd Convention, Berlin , E-Brief 310 (May 2017).
5.   G. Kidd and C. Conroy, Auditory informational masking, J. Acoust. Soc. Am. 19(1), 29-36 (Spring 2023).