INVESTIGATIONS INTO PROBLEMS WITH SOUND QUALITY DURING VIDEO CONFERENCES IN CORPORATE MEETING ROOMS

Glenn Leembruggen — Acoustic Directions, ICE Design Australia

Mark Hanson — Hanson Associates, ICE Design Australia

Camille Hanrahan-Tan — Acoustic Directions

Mark Thompson — Hanson Associates, ICE Design Australia

1 INTRODUCTION

1.1 The Problem

With the huge uptake in remote conferencing, it is vital that speech clarity and acoustic comfort for listeners is considered in the design of conferencing systems. With the advent of beam-steered microphones and the signal processing applied by platforms such as Zoom and Teams, it is tempting to think that this new technology has solved the problems. However, this is not the case, and the clarity and naturalness of speech in conferences is often poor with the result that words are missed, and concentration wanes due to increased listening effort.

Our team was asked by a large investment house to improve the quality of their listening experience and privacy in meetings with remote participants, which took place in a number of rooms. To improve the listening experience requires much more than simply improving the speech intelligibility as measured by the speech transmission index. Many other factors need to be addressed.

This paper explores the degradation in sound quality from the talker to the remote listener in corporate environments in the acoustic, electro-acoustic and electronic domains. It does not deal with the speech privacy problems.

1.2 Components in the transmission chain

To improve the listening experience of meeting participants, it is necessary to approach the problem holistically, by understanding and addressing each component in the transmission chain between the talker and remote listener.

Figure 1 shows the components in the transmission chain of a corporate remote meeting situation.

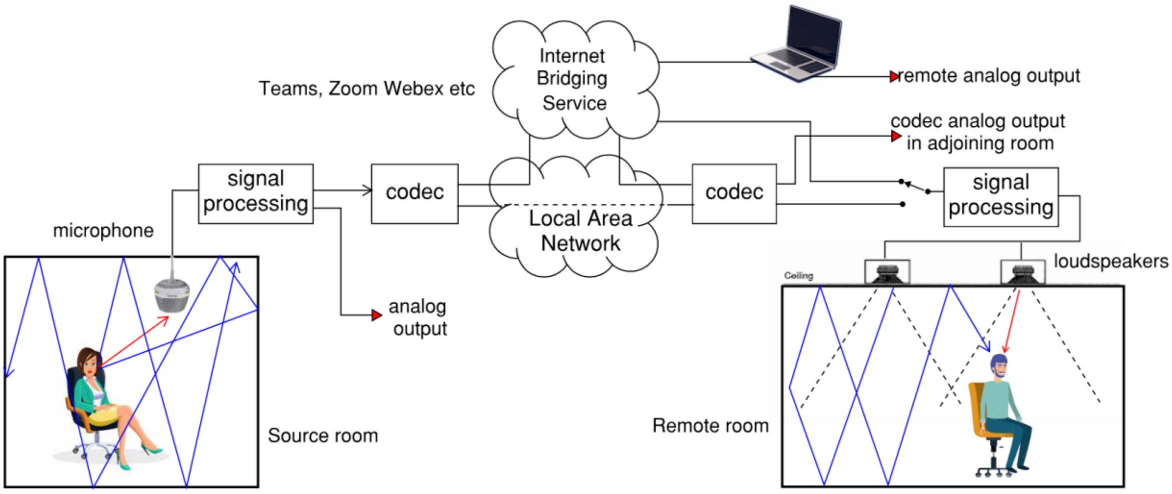

Figure 1. Components in the transmission chain of a corporate remote meeting ituation.

1.3 Our Listening Experience

Using our client's preferred meeting system Microsoft Teams, test calls were made from the source room to a remote listener in our office. Our listening experience to the transmission chain during that meeting revealed the audible artefacts listed in **Table 1**.

Table 1. Audible artefacts of the transmission chain during Teams meeting.

Attribute at Remote End	Technical Parameter	Probable Cause
Strong reverberance	Reverberation	Inadequate sound absorption in the source room
		Incorrect setup on the beam-formed microphone(s)
Strong colourations at low-mid and high-mid frequencies	Frequency response and reverberation time	Reverberation time that is not consistent with frequency in the source room
		Poor set-up or configuration of the digital signal processing parameters
		Incorrect setup on the beam-formed microphone(s)
High background noise with talker	Sound level of ambient noise in source room	Excessive ambient noise from the air-handling system
Strong graininess to the speech	Undefined	Codec and meeting software signal processing

2 REVERBERANCE

It is common knowledge among electroacoustic practitioners that the presence of reverberation in speech can degrade intelligibility, and the stronger and/or longer that the reverberation persists in the room, the more the intelligibility is degraded.

Meeting rooms are simple situations, so surely it must be straightforward to specify acoustic treatment for these spaces to control reverberation. What can go wrong if you use the Sabine equation?

The problem commences with the failure to think about the way in which sound moves in small rooms and is exacerbated by aesthetic designers who do not understand the importance of acoustic comfort. Factors contributing to this situation are:

- If the room is not acoustically diffuse, (which most are not), reverberation will occur in three discrete planes, forward-to-back, side-to-side and vertically. Each of these planes has their own reverberation time [1]. The Arau-Puchades equation predicts reverberation times when the room is a rectilinear cuboid with sound absorption that is unevenly distributed over the three planes.
- In rooms with parallel walls, room modes, flutter echoes and reverberation in one plane can manifest in similar ways, although room modes are most obvious at lower frequencies. Flutter echoes and single-plane reverberation can visually merge together depending on the arrival time of each reflection.
- Strong room modes can also occur between hard ceilings and large meeting tables.
- Currently, interior designers seem obsessed with hard surfaces and "clean lines".
- There is an understandable desire for as much glazing as possible in some rooms.
- Ceiling tiles are often the only sound absorption in meeting rooms, with the walls being acoustically hard and therefore reflective. While this may seem the correct way to control reverberation, this allows horizontal reverberation to dominate.

2.1 Example of a Meeting Room

One of the rooms in the investment-firm's offices that we worked extensively in was a twelve-person meeting room.

The room has extensive glazing, timber and plasterboard panels. The floor is covered with low-pile carpet, and the inner section of the ceiling is perforated plasterboard with an acoustic-textile backing. **Figure 2** shows images of the room, which overlooks Sydney Harbour. A large meeting table dominates the room.

Figure 2. View of meeting room.

2.2 Measured Reverberation Times

Impulse responses (IR) of the room were captured using a Type 1 microphone and WinMLS 2004 software with balloons and an NTI TalkBox as the sound sources. Seven balloon bursts and three sine-sweeps were used as the signal sources.

The reverberation times were computed from the Schroeder decay plots over the -5 dB to -30 dB decay range, with the averages compared in **Figure 3**. The agreement is pleasing, given that the directional properties of the two source types are vastly different.

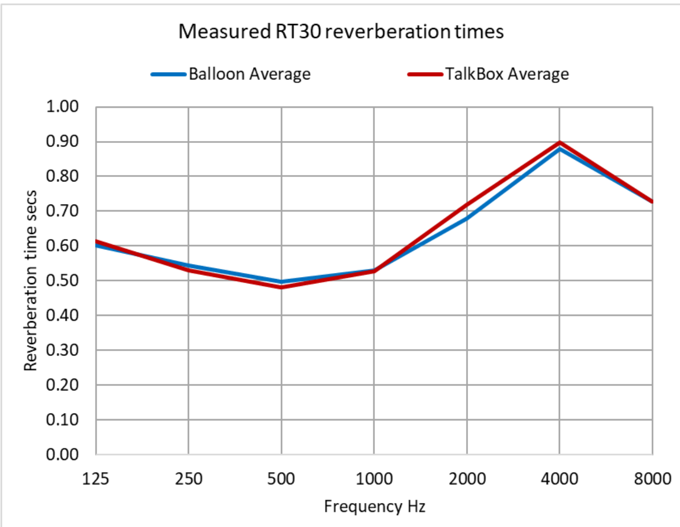

Figure 3. Measured average reverberation times (T30) for balloon bursts and with TalkBox.

From those impulse responses, the scalograms of the decays were computed using wavelet transforms. **Figure 4** shows an example of the scalogram of an IR obtained with the TalkBox while **Figure 5** shows the scalogram obtained with a balloon.

Features of the scalograms are:

- Flutter echoes in the range 3 to 10 kHz, evident primarily in the TalkBox IR.
- Protracted decay in the 3 kHz to 5 kHz range, particularly in the balloon burst.
- The more omnidirectional-like radiation of the balloon burst is evident with strong arrivals evident much later than the TalkBox.
- Room modes at 133 Hz and 195Hz, which are clearer in the TalkBox IR.

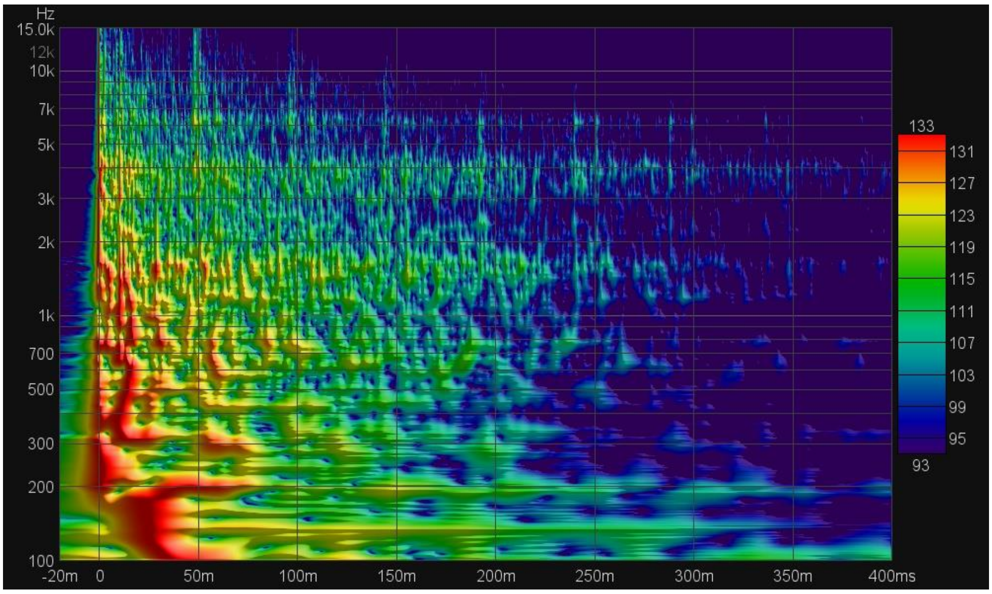

Figure 4. Wavelet scalogram of impulse response with TalkBox.

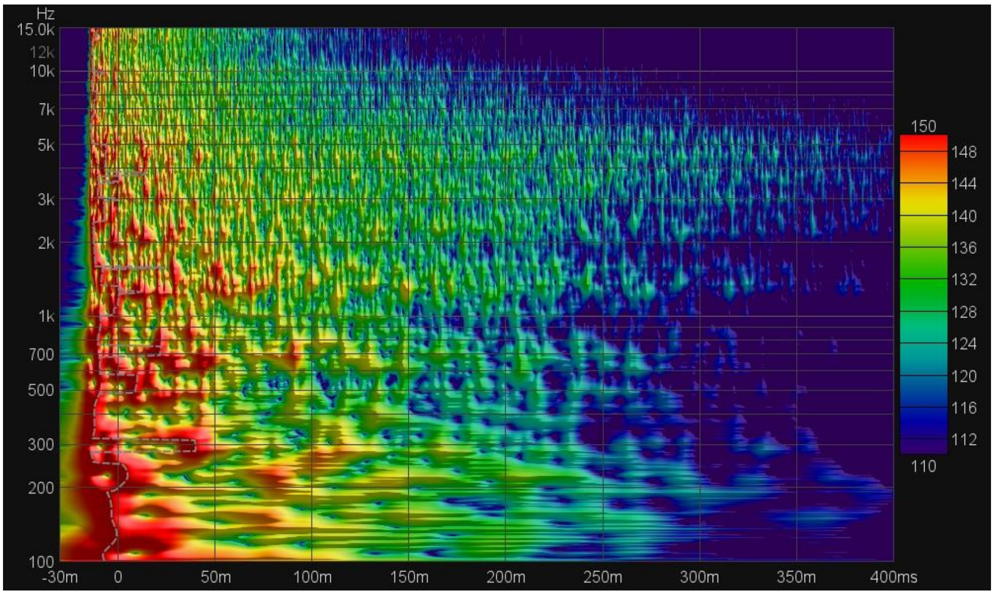

Figure 5. Wavelet scalogram of impulse response with a balloon.

2.3 Predicted Reverberation Times

Using sound absorption data for the room surface finishes and the room dimensions, we predicted the reverberation times (RTs) for the room using the Sabine, Eyring and Arau-Puchades (AP) equations.

The AP equation predicted a large boost in RT centred at 8 kHz due to an absence of sound absorption in the horizontal plane. However, its predicted values were too high at higher frequencies, and we hypothesised that the large table in the room was somewhat “trapping” the sound at these frequencies, reducing the strength of successive reflections between the walls. To make the predicted values match the measured values, the additional absorptions shown in **Table 2** were added by trial and error.

Table 2. Sound absorptions and areas added to the room surfaces to make the RT predictions using the AP equation match the measured RTs.

	Area m^2	Octave Band Centre Frequency							
		63	125	250	500	1000	2000	4000	8000
Front wall	2				0.50	0.50	0.50	0.50	1.00
Side wall 1	2				0.2	0.50	0.2		1.00
Side wall 2	2					0.50			1.00

Figure 6 compares the predicted RTs using the three equations with those measured. It is clear that the Arau-Puchades equation provides a superior match, although some minor adjustments were required to account for diffusion and “trapping” by the table.

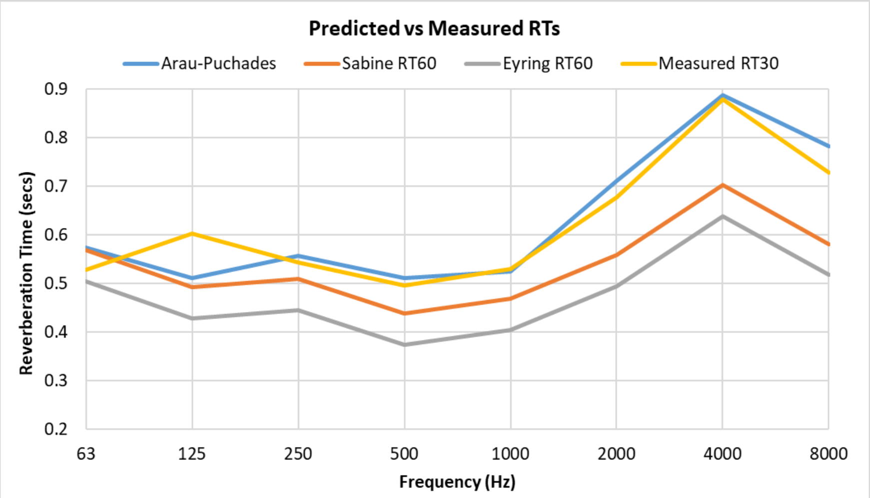

Figure 6. Comparison of predicted and measured reverberation times.

2.4 Discussion

The following conclusions are made:

a) The meeting room in which we made the measurements was typical of many Australian meeting rooms. This room showed considerable non-diffuse temporal behaviour, due to protracted reverberation time at high frequencies, some flutter echoes and room modes. Each of these attributes was audible both in the room and in the far-end audio.

b) With careful consideration of the obstruction of horizontally-travelling reflections effectively creating some additional sound absorption, the Arau-Puchades equation is likely to be the most reliable predictor of reverberation times in rooms without a high degree of diffusion. Meeting rooms are typical of these types of rooms.

c) Sources with different directional properties elicit slightly different temporal behaviours in the measured impulse responses.

3 FREQUENCY RESPONSE

3.1 Importance of Frequency Response

In many situations, the speech intelligibility of the transmission channel is often measured using the speech transmission index (STI) [2] [3]. However, investigation by an author et al [4] [5] [6] and notes in [3] indicate that STI is relatively blind to the effects of poor frequency response when the sound source is amplified. The authors consider that the loss of intelligibility in these situations is primarily due to self-speech masking in the human ear, rather than masking by introduced noise in the signal chain.

Given that our perceptions of the far-end speech intelligibility being degraded by poor frequency responses, we undertook a number of measurements to understand the scale of the changes.

3.2 Measurements

The echo-cancelling system in the source-signal processing and the encoding algorithms in the room codec and meeting software are strongly sensitive to most stationary test signals such as sine waves and broadband noise, and these devices remove those signals from the transmission channel.

This removal precludes the use of test signals such as swept sine wave and pink noise to measure frequency response. As, the only signal that appears to be passed by these devices is speech, we elected to measure the frequency response of the transmission chain by comparing the spectra of speech at various points in the chain. A speech recording issued by Bang and Olufsen was used as the speech input in all tests.

Recordings were made of the speech signal at various points in the chain in .wav format with 24 bit 48 kHz format. From those recordings, the power spectral density of each signal was computed using the Welch method in MATLAB and then smoothed over a running one-sixth octave bandwidth. Note that these spectra are different to those which would be measured in a standard one-sixth octave analyser which integrates the energy in each band.

Figure 7 to **Figure 16** show examples of the changes in spectra (i.e., a frequency response change) between specific points in the signal chain within specific rooms in our client’s Melbourne office and our Sydney office. These plots are intended to illustrate the changes in frequency response which we measured and could readily occur in similar situations. Comments are made directly below each figure.

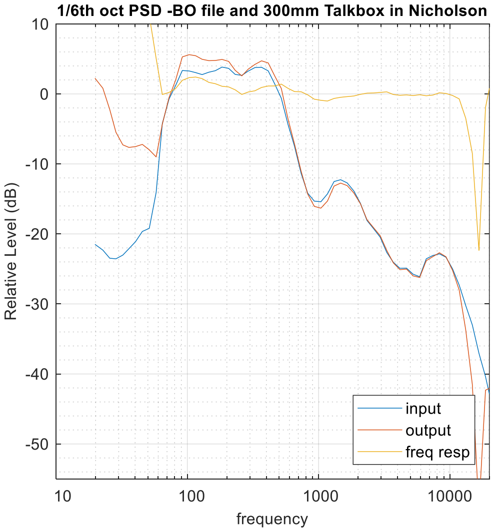

Figure 7. Response from replayed speech file to output of TalkBox at 300 mm on axis.

As expected with a device such as a TalkBox, the response at 300 mm from the loudspeaker is very similar to the input signal over the device's rated bandwidth of 100 Hz to 10 kHz.

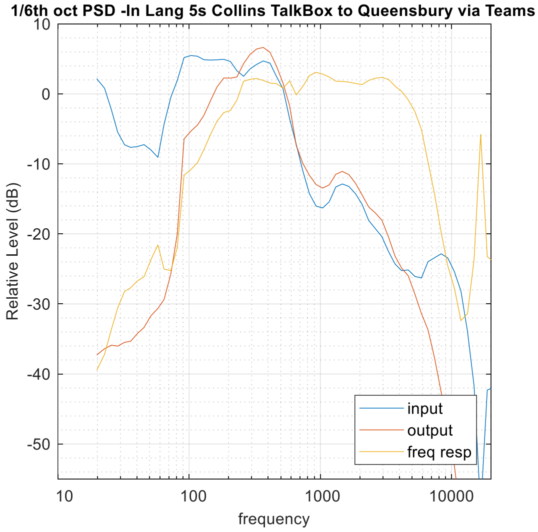

Figure 8. Collins: Response TalkBox at 300 mm to computer in adjacent room via Teams.

The system including Teams has a high pass filter (HPF) at 300 Hz and a low pass filter (LPF) at 4.5 kHz.

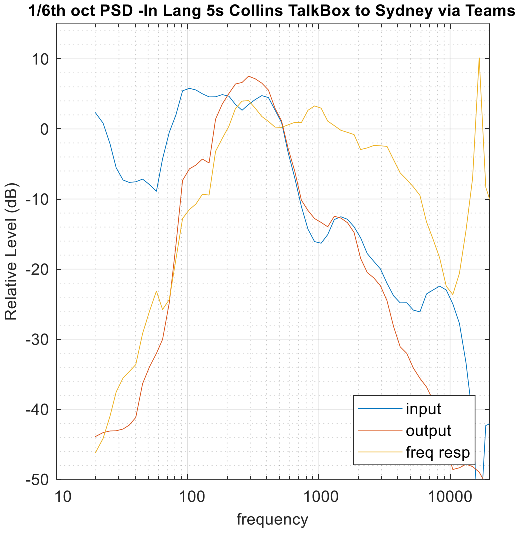

Figure 9. Collins: Response -TalkBox at 300 mm to Sydney computer via Teams.

Compared to the codec ouput in **Figure 10**, the Teams' frequency response shows a singificant loss at high frequencies.

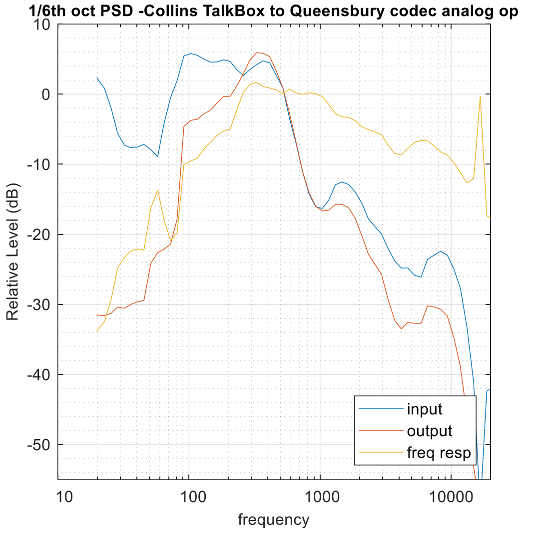

Figure 10. Collins: Response - TalkBox at 300 mm to analog output of codec in adjoining room.

With the TalkBox in the Collins room, the analog codec output in an adjoining room, (which is used to drive the in-room loudspeakers) shows a gradual LPF at 1.6 kHz and a HPF at 300 Hz.

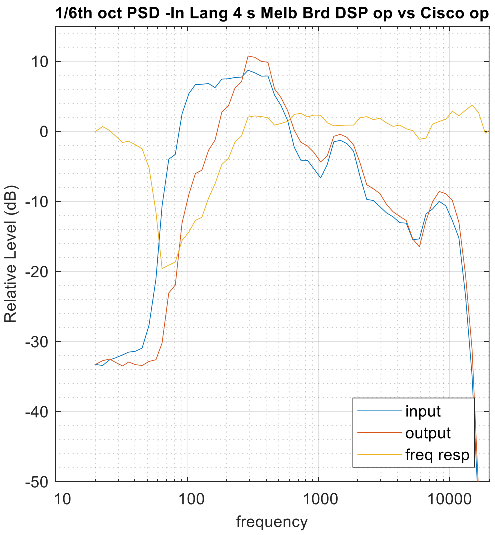

Figure 11. Boardroom: Response - analog output of DSP to analog output of codec in adjacent room.

In the boardroom, the analog codec output in an adjoining room, (which is used to drive the in-room loudspeakers) an HPF at 300 Hz and flat frequency response extending to 10 kHz.

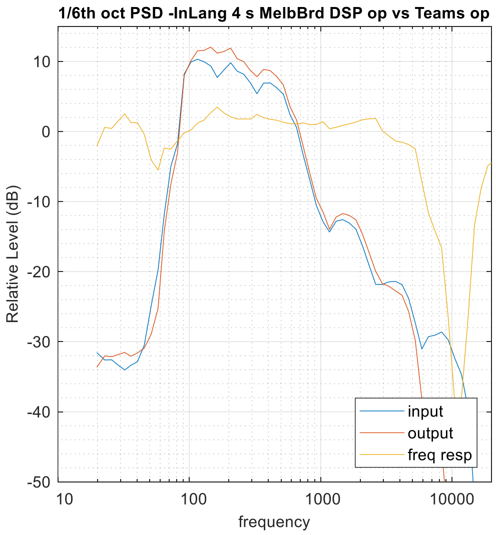

Figure 12 Boardroom: Response - analog output of DSP to remote computer via Teams.

The Teams system associated with the boardroom system has a very steep LPF at 5 kHz. Note that the peak in the response at 20 kHz is due to a “monitoring” tone at 20 kHz that is used by one of the meeting platforms.

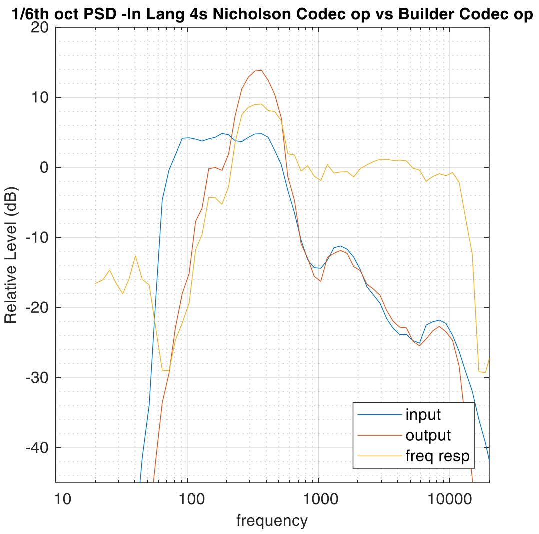

Figure 13. Nicholson: Response – dry speech to analog output of codec in adjoining room.

A peaking filter boost of 8 dB at 400 Hz from the input signal to the TalkBox to the codec output, which is used to drive the in-room loudspeakers. This boost has probably resulted from the reverberation and the DSP behaviour.

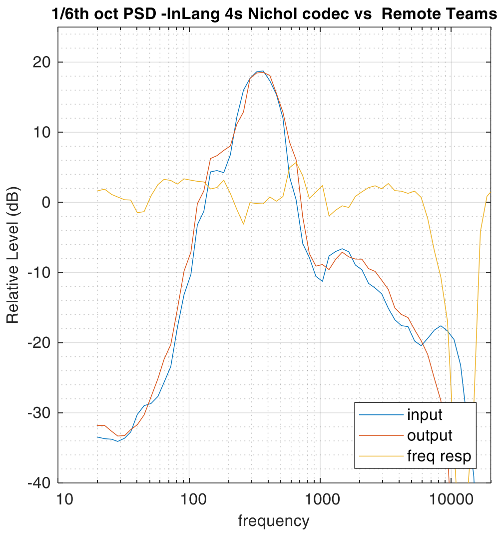

Figure 14. Nicholson: Response - analog output of codec in adjacent room to computer output via Teams.

The Teams system associated with the Nicholson room system has a very steep LPF at 5 kHz. Note that the peak in the response at 20 kHz is due to a “monitoring” tone at 20 kHz that is used by one of the meeting platforms.

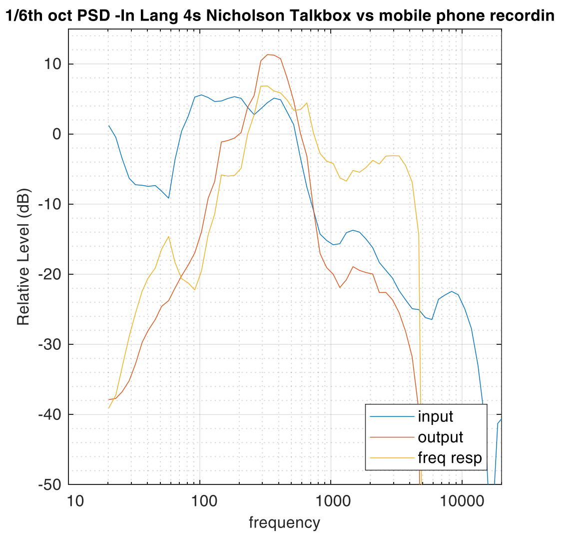

Figure 15. Nicholson: Response - output of TalkBox to mobile phone recording.

In addition to the 8 dB peak noted for **Figure 13**, the mobile phone shows a steep LPF at 3 kHz, which is expected for Global System for Mobile (GSM) transmission.

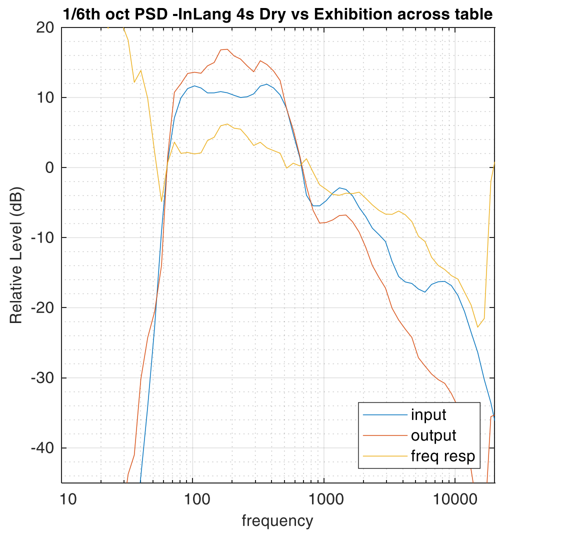

Figure 16. Exhibition room: Response - recorded speech via TalkBox to microphone at typical seat across table.

This is a good example of room reverberation damaging the frequency response.

4 NON-LINEAR DEGRADATION FROM TRANSMISSION SYSTEM

4.1 Quantifying the audio attributes of codecs and meeting software

As noted above, video-conference audio-transmission systems removes test signals from the transmission channel that are often used to quantity the audible artifacts introduced by a signal chain.

We therefore elected to explore ways to measure the degradation in quality of the speech and the use of coherence came to mind.

4.2 Definition of Coherence

Coherence (commonly known as magnitude-squared coherence) is function versus frequency that indicates how much of the output power is linearly related to the input power in the system. As such, it is an indicator of the quality of the frequency response function (FRF). Coherence evaluates the consistency of the FRF over a number of calculations called cycles. The coherence function measures the linear interaction between any two time-series in the frequency domain and has value between 0 and 1, where 0 indicates no linear relationship, and 1 indicates a perfect linear relationship.

The noncoherent remainder can be caused by the following effects:

a) Nonlinearity in the device under test

b) Noise in the device or measurement system

c) Jitter in input or output

d) Different sample rate and/or clock rates

e) Uncompensated delay between input and output

f) Analysis leakage due to an inappropriate time window or insufficient measurement resolution

g) Insufficient averaging

h) Leakage from other inputs or outputs

Effects a) through c) are caused by the device under test. Effects d) through h) are characteristics of the measurement that can be minimized by proper setup.

The coherence between two signals x(t) and y(t) at frequency f is a real-valued function that is defined as:

$$C_{xy}(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)} \tag{1}$$

where $G_{xy}(f)$ is the cross-spectral density between signals x and y and represents the correlation between the two signals at that frequency. $G_{xx}(f)$ and $G_{yy}(f)$ are the auto spectral densities of x and y respectively at frequency f and are actually power spectra.

Using coherence as a metric has a number of hazards and in one instance is a misnomer. For example, if signal x is a speech signal and signal y is pink noise, and only one FFT cycle is used to calculate $G_{xy}(f)$, $G_{xx}(f)$ and $G_{yy}(f)$, the coherence is 1. Given that pink noise is nothing like speech, this result does not equate with the intuitive perception that as speech and pink noise have little in common the coherence should be poor.

The answer to this conundrum lies in that a number of windowed FFT cycles are used to calculate coherence of a signal.

There is little published research that attempts to correlate coherence scores with other acoustic metrics that attempt to describe clarity and intelligibility of speech signals. We felt that it was important to start the process of making these correlations so that others could usefully employ the coherence metric in their acoustic signal transmission quality assessments.

4.3 Exploring the Behaviour of Coherence

We made recordings of speech and various points in the transmission chain and compared the coherence between two sections of that chain. Some of the results were unexpected and given that we were unfamiliar with coherence measurements, we elected to explore the behaviour of coherence calculations with different signals.

Coherence was computed using the function *mscohere* in MATLAB software, which estimates the magnitude-squared coherence function using Welch's overlapped and averaged periodogram method.

Table 4 below describes the parameter being explored and our conclusions. and indicates the relevant figures with coherence plots.

It should be noted that:

a) When speech is the input signal, it is a dry speech with length between 5 s and 27 s. The output signal is an altered version of the input signal.

b) To allow the trends to be more readily seen, the coherence plots are smoothed over a one-twelfth octave range.

c) Note that unless indicated, all speech files used for coherence calculations have been time-aligned within 5 ms.

d) In each of the coherence plots, the ratio of the window length to the overall length of the signal-pair is stated as: Window Length/File Length. The reciprocal of this ratio indicates the number of window cycles that were used to calculate coherence for each file length. The ratio is called R is this paper.

e) **Table 3** below lists the relationships between window order and window lengths (WL) at a sample rate of 48 kHz.

Table 3. Relationship between window order and window length (WL) at sample rate of 48 kHz.

Window Order	Window Length
13th order	171 ms
14th order	341 ms
15th order	682 ms
16th order	1.36 s
17th order	2.73 s
18th order	5.46 s

Table 4. Details of experiments with coherence and conclusions.

Parameter Explored	**Conclusion**	**Figure Number**
Effect of window length on coherence with completely different signals; speech and pink noise. Signal length was 27 seconds.	Coherence is 3% for R=5% 11% for R=19%	**Figure 17**
Effect of filtering dry speech with 1 kHz 12th order infinite impulse response (IIR) low-pass filter	Linear filtering does not change coherence. 100% coherence up to the point of 3 x the filter cutoff frequency at 60 dB attenuation.	**Figure 18**
Coherence of two decorrelated pink noise signals	Coherence is 4% for R=7% 8% for R=14%	**Figure 19**
Effect of reverberation (1 s approx.) on dry speech.	With R=1% (84 cycles), the coherence is 80% to 95%. With R=5%, the coherence is 95%.	**Figure 20**
Effect of delays of 30 ms and 60 m applied to dry speech	Averages of 92% with 30 ms delay and 68% with 60 m delay.	**Figure 21**
Effect of pink noise added to dry speech at signal to noise ratios of -10 dB and – 20 dB computed on Leq basis.	Coherence is quite sensitive to noise.	**Figure 22**
Effect of adding reverberation (RT= 1 s) to the speech-with-noise signal at SNR of -10 dB;	Increase in coherence between 14th to 16th order is approximately 10%. Comparison with **Figure 22** shows a minor decrease in coherence when reverberation is added to the noisy speech signal.	**Figure 23**
Effect of adding reverberation (RT= 1 s) to speech-with noise signal at SNR of -20 dB;	15% increase in coherence from 14th to 16th order window length.	**Figure 24**

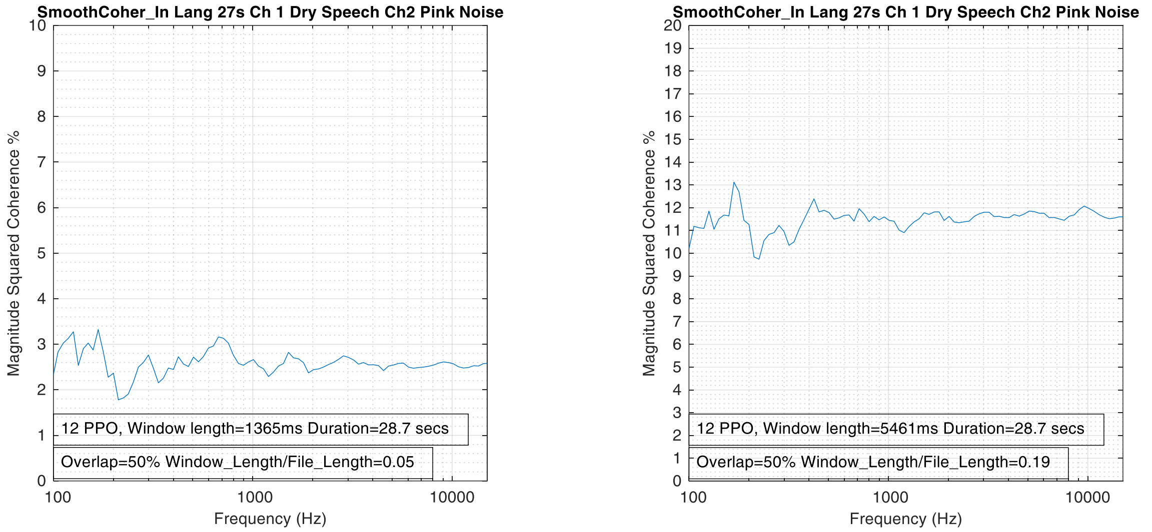

Figure 17. A left - i/p dry speech vs pink noise left 16th Ord-WL, **B** right- 18th Ord-WL

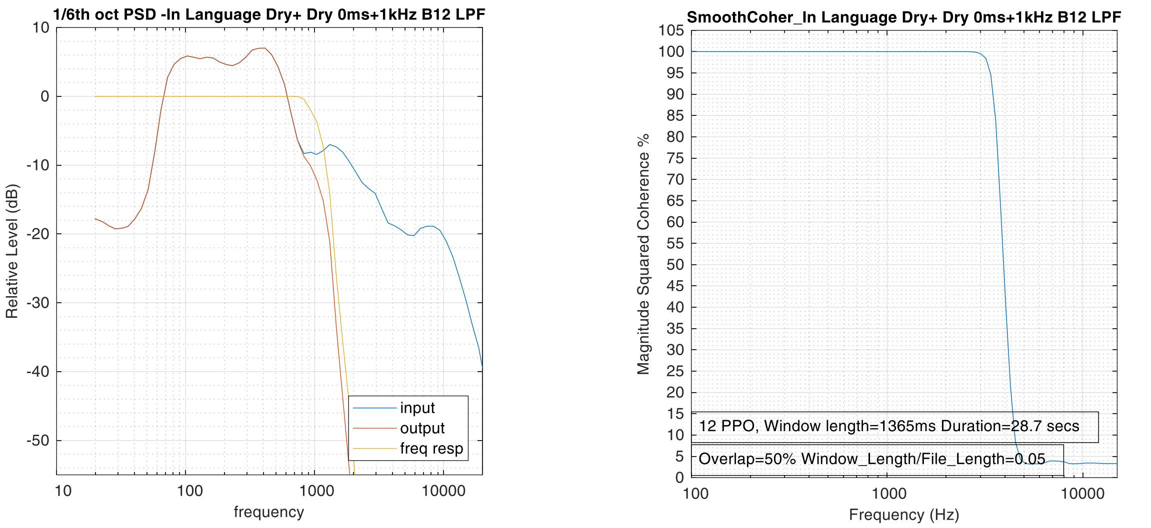

Figure 18. Dry speech vs dry speech with 1 kHz LPF **A l**eft: Frequency response, **B:** right. Coherence

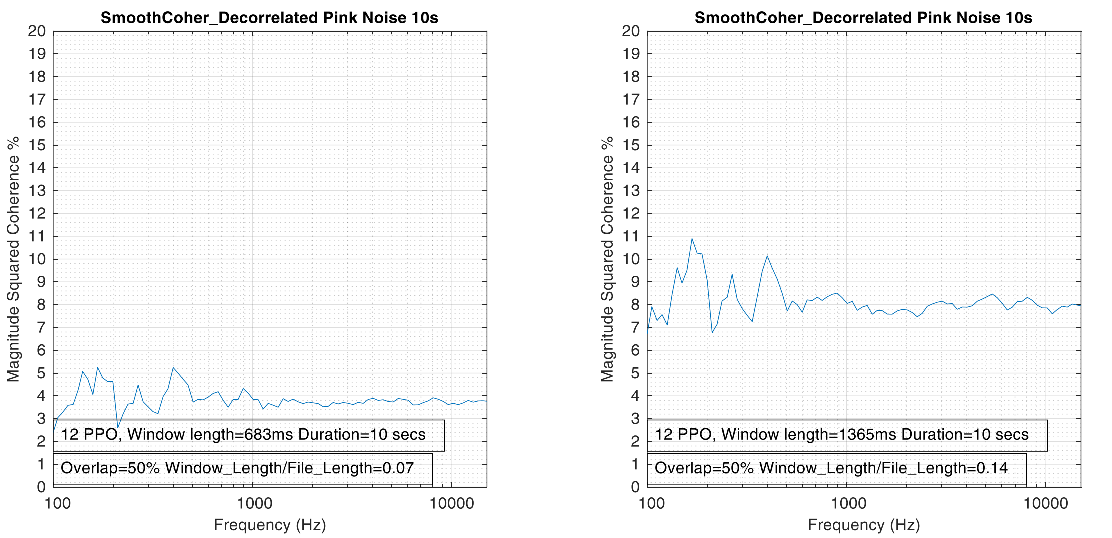

Figure 19. Decorrelated pink noise on both channels. **A** 15th Ord-WL, **B**. Right 16th Ord-WL

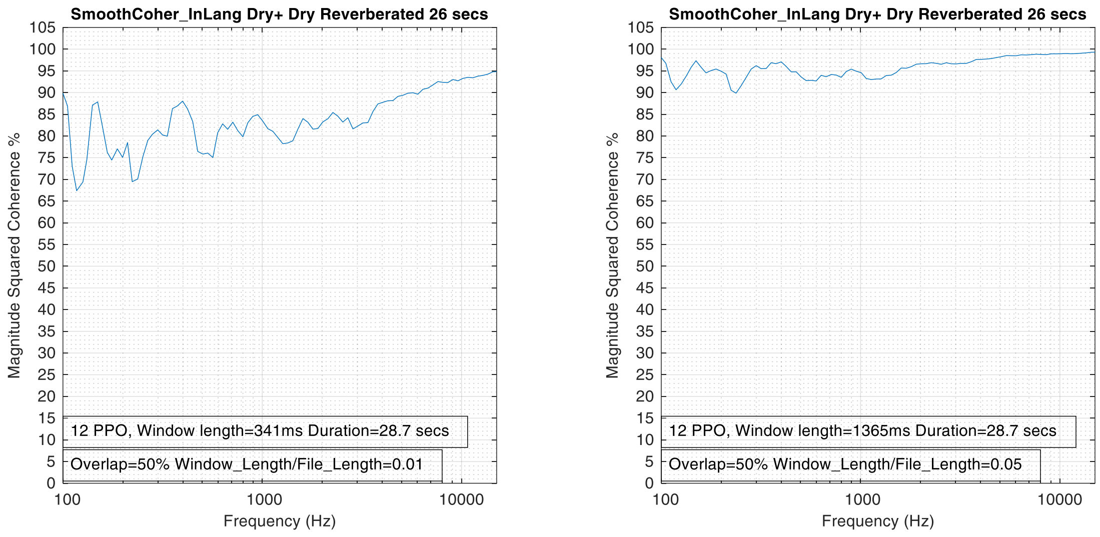

Figure 20. Dry speech with reverberation **A** left 14th Ord-WL, **B** right 16th Ord-WL

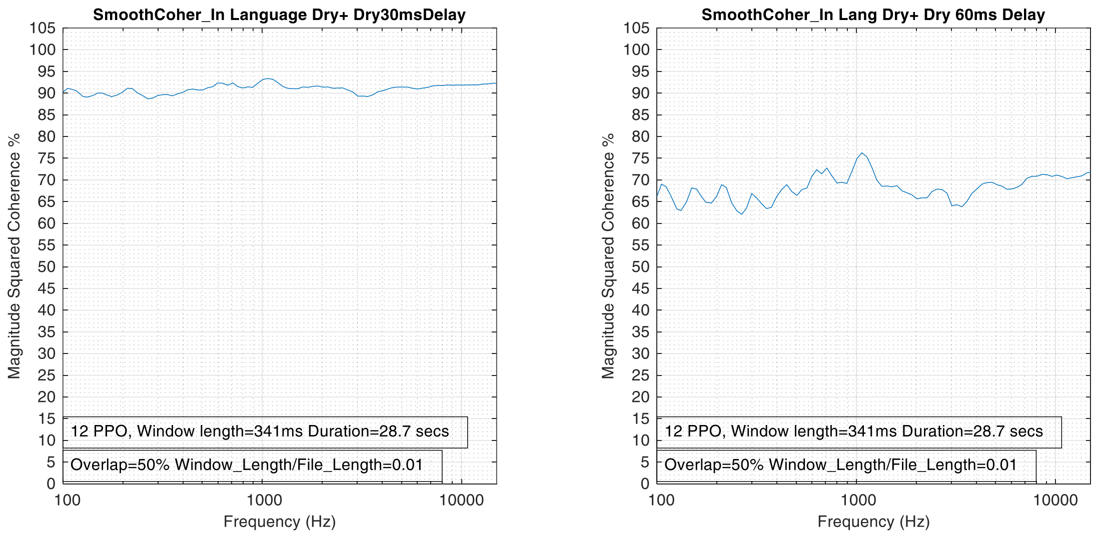

Figure 21. Dry speech with delay added.14th Ord-WL **A** left 30 ms delay, **B** right 60 ms delay

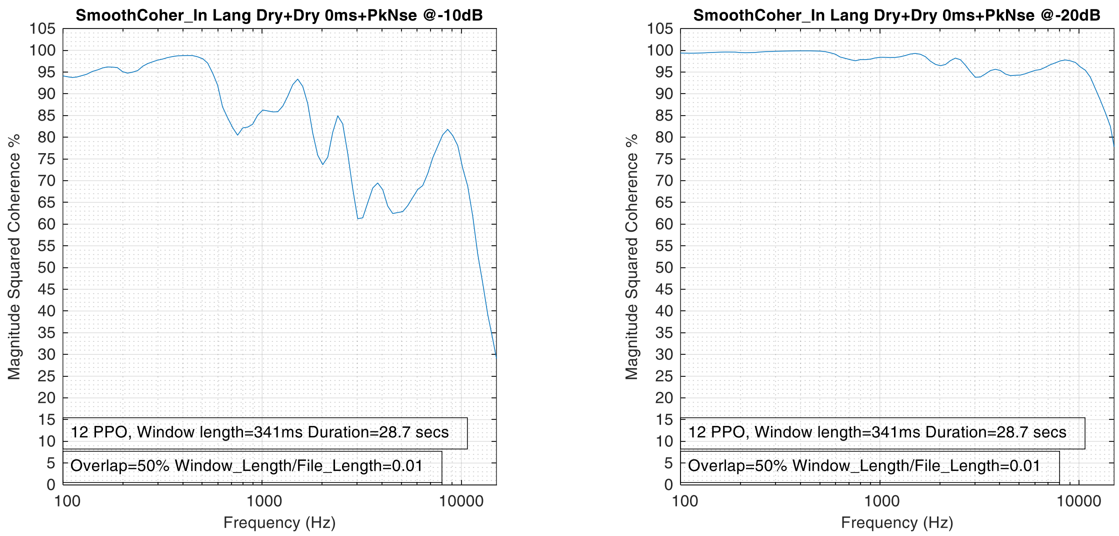

Figure 22. Dry speech with noise added. 14th Ord-WL **A)** SNR = -10 dB, **B)** SNR = -20 dB

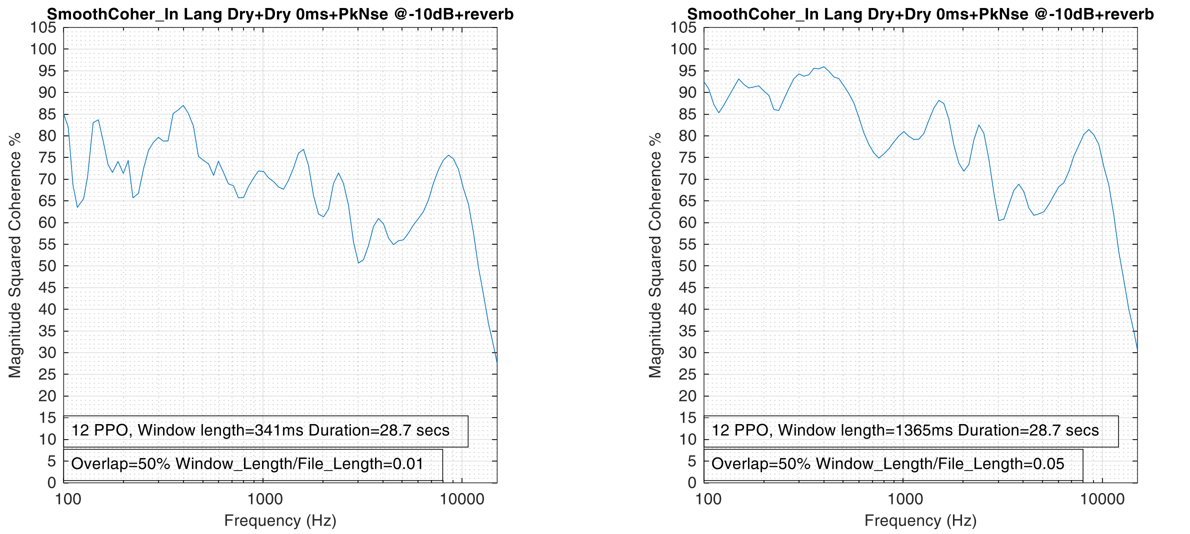

Figure 23. Dry speech with pink noise SNR -10 dB and reverberation. **A**) 14th Ord-WL **B**) 16th WL

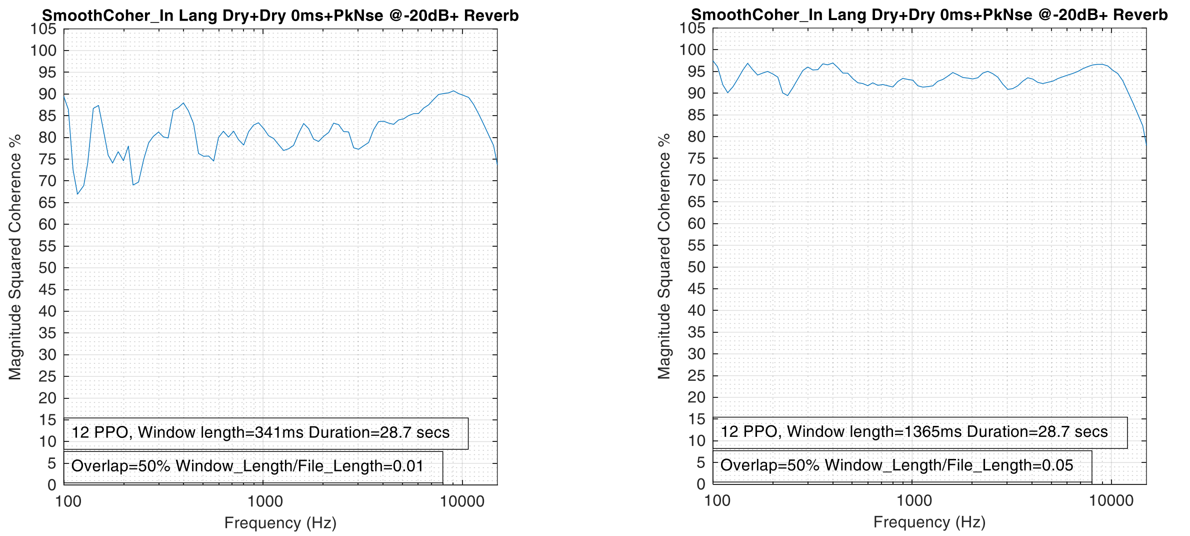

Figure 24. Dry speech with pink noise SNR -20 dB and reverberation. **A**) 14th WL **B**) 16th WL

4.4 Discussion of Experiments with Coherence

- Coherence values are very sensitive to the length of the analysis window and the duration of the file being analysed.
- Coherence is relatively sensitive to small temporal offsets of the order of 10 ms between the two signals. For example, a 30 ms offset produces an average coherence of 90% whereas a 60 ms offset produces an average coherence of 68%.
- Coherence is blind to frequency response at the point at which noise in the signal becomes significant. This is borne out in **Figure 18** B.
- The longer the analysis window, the higher the coherence.

4.5 Coherence in the Meeting Room Situation

Coherence plots were made for a range of example situations in the meeting rooms, with the list and associated figure shown in **Table 5**. As the coherence values in the room situations were quite low with long speech segments and small window lengths (WL), we elected to use a speech segment length of 4.5 seconds with a window length of 683 ms (15th order). The Ch1 and Ch2 signals were visually time-aligned to within 3 ms, which was confirmed by cross-correlation calculations.

Table 5. Examples of loss of coherence (all 15th Order WL with 4.5 s long speech).

Input signal (Ch1) in coherence calculation	**Output signal (Ch2) in coherence calculation**	**Figure**
Speech recording file^ broadcast in room	300 mm from TalkBox on-axis.	Figure 25
	Seated position in Latrobe room due to reverberation, noise and TalkBox – strongly reverberant.	Figure 26
	Seated position in Exhibition room due to reverberation, noise and TalkBox.	Figure 27
Analog output of codec in adjacent room	Exhibition room – to remote computer via Teams.	Figure 28
TalkBox output in Collins room	Analog output of codec output in adjacent room.	Figure 29
	Remote computer in Sydney via Teams.	Figure 30
Speech recording file^ broadcast in room	Analog output of DSP in Boardroom.	Figure 31
Analog output of DSP in Boardroom	Analog output of codec op in adjacent room.	Figure 32
	to remote computer via Teams.	Figure 33
Speech recording file^ broadcast in Nicholson room	Mobile phone recording.	Figure 34

^ anechoic recording from Bang and Olufsen

Table 6 holds the figures listed in **Table 5** and provides some commentary.

Table 6. Coherence results and commentary.

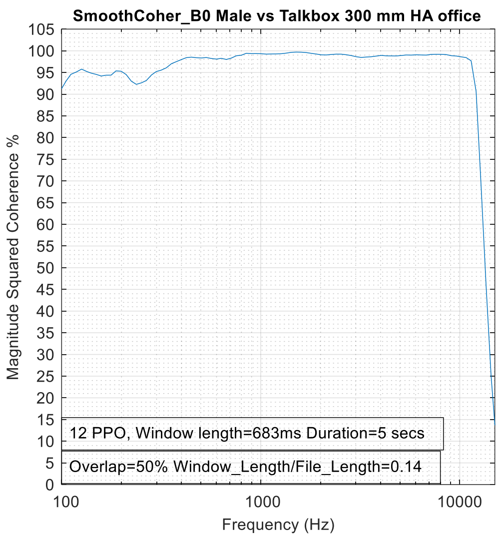

Figure 25. Coherence: At 300 mm from TalkBox with raw (dry) speech.

As expected from a high-quality loudspeaker, the coherence is high, almost 100%. Sounds virtually identical to the recording.

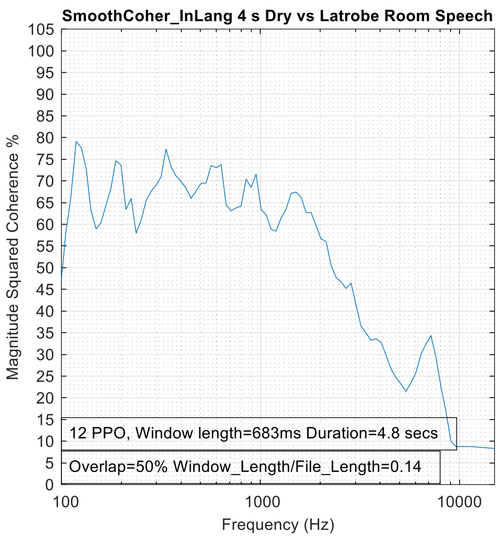

Figure 26. Coherence: At seat in Latrobe room from reverberation, noise, and TalkBox.

Listening shows high reverberance with strong boost at midrange frequencies.

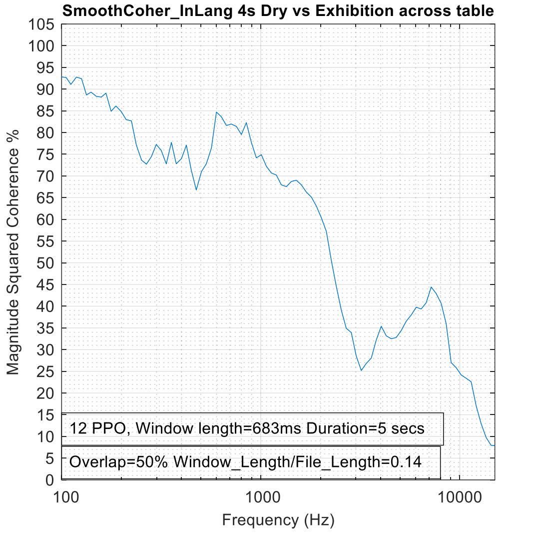

Figure 27. Coherence: Dry speech to seat in Exhibition room.

Listening reveals considerable reverberance with reduced high frequency content.

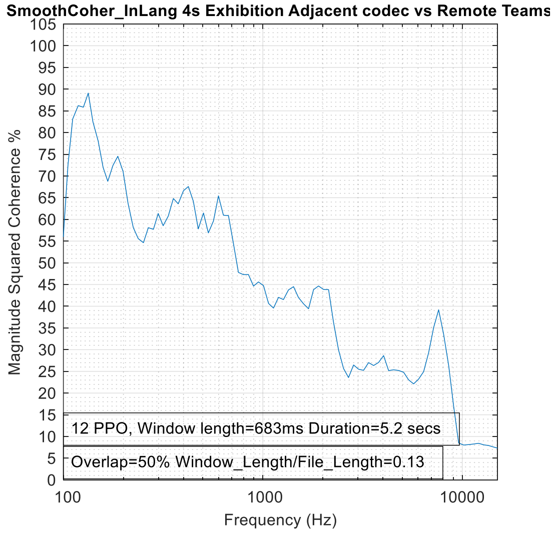

Figure 28. Coherence: Exhibition room - analog output of codec in adjacent room to remote computer via Teams.

Listening to both the codec and Teams outputs shows considerable reverberance. Teams has less low frequency content. Given that both inputs to the coherence calculation are reverberant from the same source, it is surprising that the coherence is so poor.

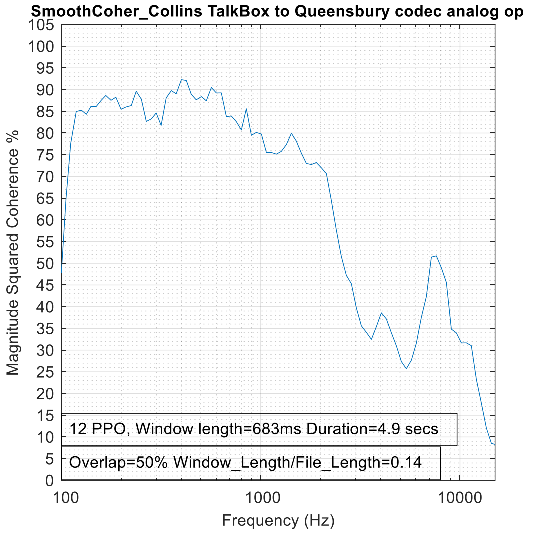

Figure 29. Coherence: TalkBox output in Collins room to analog codec output in adjacent room.

At the local codec output, the coherence falls from 85% to 70% at 2 kHz, after which it rolls off (i.e., decreases) steeply.

Listening indicates moderated reverberance with boost at midrange frequencies.

Some graininess to the codec output.

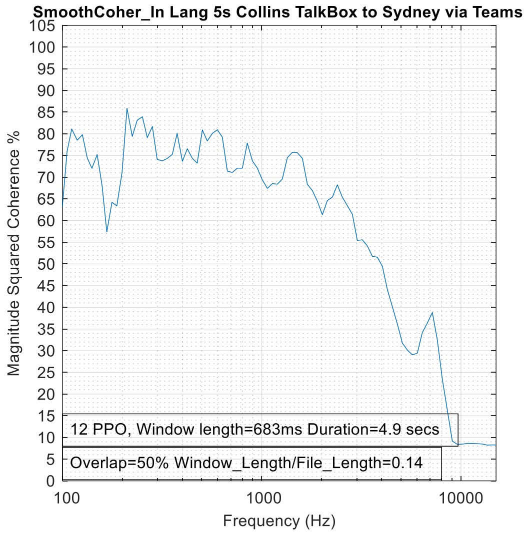

Figure 30. Coherence: TalkBox output in Collins room to remote computer via Teams.

However, at the remote computer, the coherence is approximately 10% lower than at the codec up to 2 kHz, but the roll off is much more gradual.

Listening indicates moderated reverberance with boost at midrange frequencies with less low and high frequency content that at the codec in **Figure 29**. Some graininess to the Teams sound.

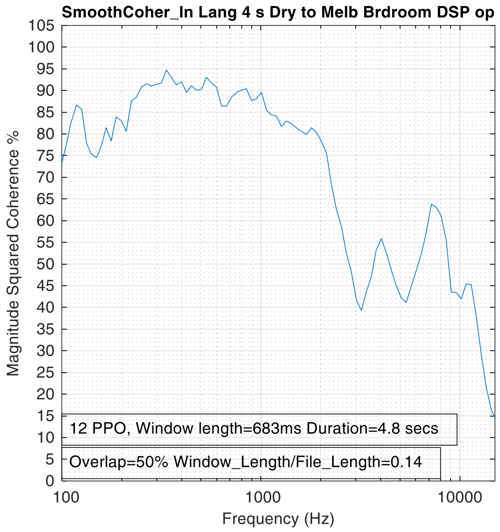

Figure 31. Coherence: speech recording to analog output of Boardroom.

Tonal variations are evident in the recording. Minor reverberance is present.

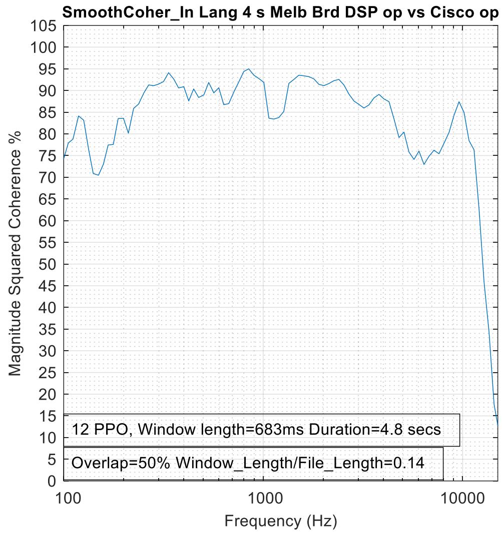

Figure 32. Coherence: analog output of Boardroom DSP to analog op of codec in adj. room.

Listening reveals considerably less low frequency content than at the DSP output. Some graininess to the codec output.

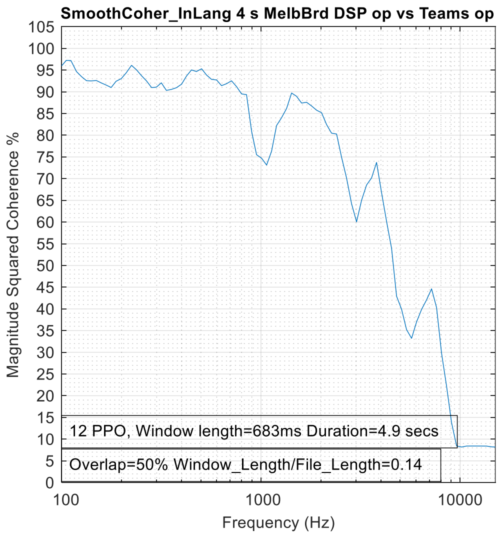

Figure 33. Coherence: analog output of Boardroom DSP to remote PC via Teams.

Listening reveals similar tonal content than at the DSP output. Graininess not strongly evident.

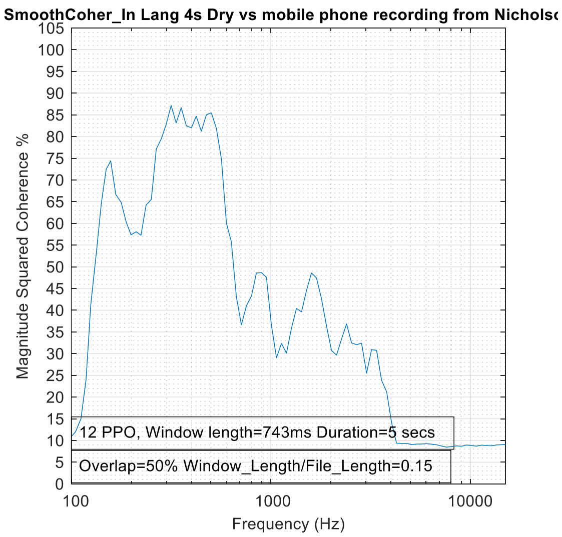

Figure 34. Coherence loss from raw speech in Nicholson room to mobile phone recording.

Graininess is hugely evident in the recording, along with absent low and high frequency content.

4.6 Discussion

After a deep-dive into the use of coherence to quantify the degradation in speech quality and listening comfort after transmission through both linear and non-linear processed systems, we are still unsure about how to use coherence as a metric. One of the most puzzling aspects is the loss of coherence at high frequencies in the meeting room situations, which was not strongly evident in the experiments.

We conclude that coherence is likely to be a useful comparison tool for various situations if its calculation parameters such as signal and window lengths are constant. However, we are unclear about setting a threshold for pass/fail situations.

Note that coherence cannot be used to assess frequency response.

5 CONCLUSION

Our exploration of the degradation in sound quality from the talker to the remote listener in corporate environments in has encompassed the acoustic, electro-acoustic and electronic domains.

In the acoustic domain, our example of a typical meeting room exhibited non-diffuse reverberant decay, flutter echoes and room modes. The best predicted match to the measured reverberation time (T30) was with the three-dimensional Arau-Puchades equation.

As quasi-stationary test signals are stripped out by remote conferencing software, speech was used to investigate the frequency responses and the loss of coherence within the transmission chains.

Computation of the spectral differences between the recorded speech file and the outputs of adjacent-room codecs, the remote Teams computer and mobile phone show important degradations in frequency response at various points in the signal chain.

As we were not well acquainted with coherence measurements, we undertook some experiments using speech and noise signals. Those results may assist readers who are interested in this area.

We also computed the coherence between various sections of the speech transmission channel in an attempt to quantify the degradation in speech quality and listening comfort. However, given the outcomes, we are still unsure about how to use coherence as a metric for speech degradation through remote conferencing systems.

A separate body of research is required to assess coherence results against extensive subjective tests so that relationships between coherence scores and perceived quality can be developed for field use.

6 REFERENCES

[1] H. Arau-Puchades, “An Improved Reverberation Formula,” *Acta Acustica united with Acustica,* vol. 65, no. 4, pp. pp. 163-180, March 1988.

[2] H. Steeneken, T. Houtgast, “Basics of the STI measuring method,” *Reprint from : Past, present and future of the Speech Transmission Index.,* https://www.steeneken.nl/wp-content/uploads/2014/04/Basics-STI_2014-V1.pdf ISBN 90-76702-02-0.

[3] *Objective rating of speech intelligibility IEC 60268-16,* International Electrotechnical Commission, 2020.

[4] Leembruggen, “Another look at the relationships between frequency response and the speech transmission index, with respect to word scores and road tunnels,” *Proc IOA,* vol. 41, no. 3, 2019.

[5] 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. 2009 .

[6] G. Leembruggen, M. Hippler, P. Mapp, “Further Investigations into Improving STI’s Recognition of the Effects of Poor Frequency Response on Subjective Intelligibility,” AES Preprint 8051, 128th Convention, May 2010.