Canarie AAP-03 "Shared Spaces" Project Milestone 1 Report
Appendix 2
Report on the Feasibility of Multichannel Echo Cancellation

Wieslaw Woszczyk, Charles Gagnon, Kent Walker

1. Executive Summary

There are no commercially available solutions for multichannel audio echo cancellation. In examining the feasibility of pursuing such a solution as part of this project, we reached the conclusion that it would require far more resources than those currently available with no assurance of success at the end of the effort.

We then tested three compromise solutions to see if they could provide adequate echo cancellation for stereo audio transmission: the commercially available Gentner 1524 "black box" device, McGill developed software that was incorporated into the "Bronto" transmission software being used in this project and a McGill developed system that uses commercially available DSP plug-ins available as part of the Metric Halo 2882+DSP (MIO). The best results were obtained using the latter MIO system although acoustical means of echo prevention will have to be added such as close microphone placement and highly directional microphones and speakers.

A detailed discussion of echo cancellation methods and the testing results follows.

2. Echo cancellation methods.

Echo cancellation is required in bidirectional communication between two sites when there is acoustic crosstalk at each site between loudspeakers and microphones, accompanied by a delay in transmission. Sound originates at a person's mouth and is monitored at the ear, therefore there is a spatial overlap between source and receiver of the sound, and hence the resulting crosstalk. When transmission delay is short, there is a quick build up of feedback loop and cannot be recognized as echo, which is a distinct and separate appearance of delayed sound when delay exceeds 50ms.

There are a few methods available for reducing the amplitude of echo, the level of delayed sound returned to the source. They typically focus on ensuring a close placement of microphones to the corresponding sound sources using narrow controlled directivity of microphones to reject the sound projected by the loudspeakers, other methods use highly absorptive acoustic environment to dampen the reflected sound paths from loudspeakers to microphones, there are also methods that use headphones and in-ear monitoring to isolate the reproduction of sound directly to the ear.

There are also methods that use signal processing to remove the echo already present in the signal, or methods that prevent echo from being created and added to the signal. Signal processing that aims at removing the echo already present suffers from latency due to block processing delays, and to conversion from time to frequency domain to perform signal analysis and digital filtering. There is also a penalty in reduced sound quality because the main signal is filtered adaptively thus creating audible artifacts. Processing methods that prevent echoes from being formed rely on digital processing in the time domain of the dynamic structure of signals without conversion to frequency domain. They are inherently without delay, or with a small delay of 1-2ms maximum due to analog to digital conversion.

Until now, McGill's methods dealt with the echo problem by using close microphone placement and their high directivity to reduce the capture of radiated sound from the loudspeakers and reflected sound from the room. There were also methods applied to fill in the time-space between the sound and its echo with components of room simulation in order to hide the echo in the continuous sound without giving it a separate identity. This current evaluation concerns the feasibility of using multichannel echo cancellation in bidirectional transmissions of music using

signal processing methods. In-ear monitoring has not been considered in the present study, although this method is viable if needed and appropriate to the application. The newest in-ear monitoring methods, introduced by Sensaphonics (http://www.sensaphonics.com/monitors.html) provide each user of molded insert-earphones with a set of in-ear microphones with which the user can create the balance between the sound present in the room, and the sound present in the sound system. Nothing from the sound monitored in the ear leaks out into the microphones. The growing acceptability of such system by musicians may become a factor in future solutions to echo cancellation.

The current study aims to evaluate if signal processing methods provide useful solutions in echo reduction (cancellation or suppression), and if so, to identify what type of processing works best and is preferred by the users and participants.

3. Multichannel, Stereo or Mono Echo Cancellation

Although our original goal was to research the suitability of using multichannel echo cancellation, we had to resign this work to stereo and mono echo cancellation. Adaptive filtering and other DSP methods providing echo cancellation are not available in multichannel configurations. Thus far, leading research companies in the field such as Fraunhofer Institute in Germany (Walter Kellerman), Philips Research (Daniel Schobben) and Lucent (Jacob Benesty) have not been able to develop a working multichannel echo canceller that works. Our tests will involve single channel and two-channel cancellers in order to provide feedback that can be used during the life of this project.

4. Systems Tested

A bidirectional audio transport was implemented to test echo cancellation schemes for audio conferencing applications. Three echo suppression systems were compared: the commercially available Gentner 1524, a DSP engine native to the packet-management system (Bronto) used for network transport of audio, and a system using native DSP plug-ins available as part of the Metric Halo 2882+DSP (MIO).

Audio was transmitted at 24bit/96kHz using the MIO as input/output on either end of the signal chain interfacing with M-Audio Delta 1010, and as a router for connection with the Gentner. Signals were routed locally (internal routing in MIO) at 24/48 digital resolution and were output analog to the transport. A/D/A conversion at both ends of transport was effected using an M-Audio 1010 interface at 24 bit 96kHz.

Duplex transmission took place between the Immersive Presence Laboratory (IPL), a research space for multisensory experience incorporating 30 channels of audio, high definition video and haptic information display, and a remote location in the same building. Transmission was across a 1 Gigabit connection to the McGill University backbone.

The Gentner 1524 is a “black box” echo canceller used for both live and teleconferencing applications. There is no information given in the system documentation about the method of echo suppression. This is a mono, single-channel system.

The Bronto system was designed by Stephen Spackman and Jeremy Cooperstock of McGill University.[1,2] The system uses proprietary software for packet-loss management as well as UDP protocol for network transport. Echo suppression DSP is also a proprietary system based on adaptive filtering.[3] Hardware consists of Dell servers running the Bronto kernel on Linux Red Hat, with audio I/O through the M-Audio 1010 PCI interface.

The Metric Halo system uses DSP available internally on the MIO 2882 and is based on dynamics control using the proprietary compression/gating plugins. We used a stereo

configuration. Fig. 23 below shows the dynamics processing used that was developed by Woszczyk.

5. Test Location

IPL is an audio-visual research lab located in the TV studio of the Instructional Multimedia Services Building of McGill. The system comprises 30 channels of audio in a 24.6 configuration. Twenty-four channels of I/O are transmitted through a crossover network to 24 four-element ribbon driver enclosures plus five LF enclosures and one motion platform. The ribbon drivers are distributed in three symmetrical (re: azimuth in horizontal plane) vertical layers of an ITU plus centre-rear derivation, approximating the inside if a semi-sphere. Subs are placed in a similar ITU distribution on the horizontal plane. (See Figs. 17 and 20 for dimensions and layout of rooms.)

Room treatment in IPL consists of heavy theatre-style curtains around the walls of the listening area, as well as bass absorbers placed to break standing waves. Ceiling treatment is of HF absorptive mousse, and corners are curved. Measured RT60 of the room is ~0.5s in the low frequency range and ~0.15 in the midrange. The remotely connected to IPL Room 1684 is located on the 16th floor if the same building, is square with recessed windows, a carpeted floor and standard office style ceiling panels, no wall treatment. There are no available RT60 data for this room.

6. System Details

Fig. 1 below shows a general schematic of the pickup and routing.

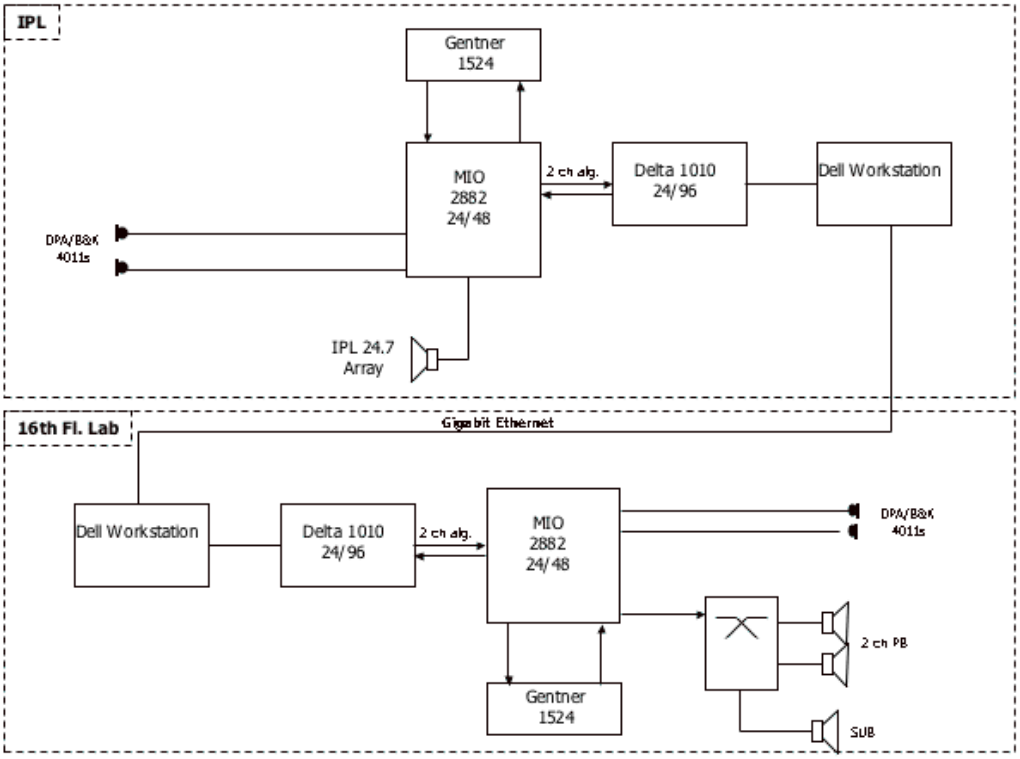

Figure 1: General signal flow.

7. Evaluation procedure
The following conditions of signal treatment were compared:

1. Gentner with echo suppression
2. Gentner with echo suppression bypassed
3. MIO DSP engaged
4. MIO DSP bypassed
5. Bronto with native echo suppression
6. Bronto with echo suppression bypassed

Audio pickup was stereo, using spaced cardioid pair (B&K/DPA 4011) of microphones for a natural-sounding reproduction of the remote space and representation of the speaker's lateral movements.[1] Fig. INS and Fig. INS show details of the signal flow. The stereo microphone was approximately 50 cm away from the mouth of the speaking person.

The noise floor in IPL was measured at 32 dB SPL, un-weighted. Noise floor in 1684 was 39 dB SPL un-weighted.

8. Objective tests
Two objective tests were performed to determine the magnitude of echo. In the first (1), 1kHz ping tone was recorded with the system on and in bypass to measure the envelope of the signal and drop in echo level. In the second (2), RASTI (rapid speech transmission index) measurements providing objective measurement of speech intelligibility were taken with simulated speaking person in front of microphones and the simulated listener standing behind (please see the photo). Since the bidirectional system introduces repeated decaying echoes, their presence will reduce the RASTI value. Thus, RASTI value predicts the loss of information (intelligibility) in the critical 500Hz and 2000Hz bands, when communication is accompanied by echo.

To determine objective echo cancellation performance a 1kHz ping tone of ~2s duration from a Brunelle 3020 signal generator was fed from the IPL center rear speaker, middle ring, and the result recorded with both 4011s and B&K HATS measurement system, for each condition of signal treatment. Signals were aligned in a DAW and echo decay envelopes were compared relative to the source ping tone.

Hardware latency was determined using a straight patch-through ping, recorded into the same multi-track file as the source, and measuring the offset in samples. This test was possible for Gentner and MIO systems only as Bronto was offline at the time.

RASTI measurements were taken using the B&K system 4225, with the HATS as pickup and the transmitter positioned in the subject position in IPL as shown in Fig. INS. Combined 500Hz and 2kHz signal was used with time integration of 32 seconds. Two sets of RASTI measurements were taken, with and without a 300Hz tone in the background to determine the influence of continuous signal being present (e.g. 3rd party) on capabilities of echo cancellation.

[1] Bronto normally incorporates video (SDI) transport in parallel with audio. Audio pickup was designed around a simulated real-world duplex videoconferencing system, with associated hardware in place and the subject positioned facing the screen as if in a live videoconferencing situation, although video transmission was not used during the evaluation.

9. Subjective tests

Two subjective tests were performed with four system users working in pairs and interacting with one another. No video was used, only audio, as we did not want any multimodal effects (lack of sync, balance, etc.) affecting the results of echo assessment.

Subjective evaluation used two sets of two subjects, asked to engage in ~5 minutes of conversation and then rate system performance on criteria of perceived signal quality (of remote speaker's voice) and perceived echo level (of one's own voice). A simple rating scale of 1-5 was employed with 5 being the highest level of quality/most perceivable and 1 being the least. Results are shown in Table 5.

10. Results

Tables 1-4 below show results from subjective evaluation tests and RASTI measurements. The sound quality of partner's voice is generally somewhat degraded when the echo reduction or suppression is engaged. In the case of MIO system this us due to pumping effect from modulation of background noise by the compression/gating process. In the case of Gentner system, this is due to degradation of the voice quality subjected to digital signal processing. Bronto system showed the lowest quality due to audible distortion and signal discontinuities. The spectral plot of the ping tone decay presented in Figure 11 compared to the reference tone in Figure 5 shows a marked difference in the case of the Bronto echo canceller. Figures 5-11 show spectrum analysis of the source ping tone followed by spectra of each system in bypassed and cancellation mode.

RASTI measurements were made on the output of in-ear microphones of the artificial listener (B&K Head and Torso Simulator, or HATS) for both the left and the right ear. The human speaker was replaced by RASTI system transmitted that generated modulated RASTI signal in front of the microphones. The decrease in RASTI value of the ear of HATS shows that echoes produced by bidirectional transmission is impeding intelligibility of speech communication. It is likely that music intelligibility will be affected in a similar way. At present there are no reliable measures established for assessing music intelligibility. The measured RASTI results for speech show the greatest improvement in intelligibility when using the MIO system (0.64 without echo cancellation, 0.90 with echo cancellation). Gentner results seem to show marginal improvement. RASTI measurements made with the 300Hz tone in the background show lower values of improvement since the 300Hz tone is narrow band and is not a part of the modulated source signal.

Figures 2-4 below show the difference in dB of the level of the decay envelope caused by the echo repeats relative to the ping tone. The left channel of the artificial listener (dummy head) in particular shows a marked difference between the two conditions: echo cancellation ON, and echo cancellation OFF. The echo level is reduced by approx. 10dB and more along the 100ms time segment for Gentner and MIO, with a visible latency of approx. 20ms to achieve maximum echo reduction. The dynamic behavior of the Bronto echo canceller seems not to be optimized, and the maximum echo reduction cannot be confirmed. In the measured results the Gentner and MIO systems performed comparably, with 0.75 dB better results from MIO. The Bronto system showed the poorest performance with 1dB difference shown for the left mic of the HATS system. However it is hard to assess conclusively the performance due to anomalies in the waveforms, likely due to time based distortion caused by the physical layout of the room. The monitoring gain was likely different as well as this system was not under control of the MIO mixer. If we take the most reliable measure to be that of HATS left, the above observations stand.

Figures 12-15 show the measured hardware latency for Gentner and MIO systems in both states, with and without echo cancellation. In the case of the MIO, measured latency was less than specified by the manufacturer, who gave a 102 sample basic A/D/A latency, with 16 samples added when DSP processing is engaged. The Gentner has no published latency values, but showed a significantly higher latency than the MIO running DSP. We should add that Bronto transport is asynchronous; round-trip system latency depends on network performance, with

anywhere from 60µs to 200ms possible. The system uses blocks of adjustable size at the buffer input. Block size for the transport during echo suppression tests was 1024 bytes. The Bronto echo suppression code adds an additional 4096 byte block to the buffer.

11. Conclusions

Subjective results showed comparable evaluation of MIO and Gentner in terms of signal quality, with the least signal quality being ascribed to the Bronto system with DSP engaged. In terms of echo audibility the MIO had the least audible echo when DSP was engaged. Since MIO system can have a number of DSP channels, and provides an integrated DSP package with up to eight microphone preamps and more digital channels, this could become the preferred option for reducing the echo in a multichannel environment involving speech and music.

It is clear that acoustics and placement precautions must be applied in order to reduce the need for extreme digital signal processing because this always produces side effects such as distortion and pumping, in addition to latency. Close microphone placement to the sources, the use of wireless microphones on person's body, the preference for properly placed directional loudspeakers and microphones, as well as dry room acoustics, can help reduce the potential for echo development. These conditions may be difficult to adhere to when many musical instruments and persons are involved.

By far the most effective method of echo cancellation is the use of in-ear monitoring that restricts the delivery of sound directly to the ears of all participants, and not into the room where microphones collect the sound. This method is used extensively by artists performing on stage where it also functions to reduce the high sound pressure level (SPL) of amplified sounds impacting the ear drums of musicians. A good example suitable for use in bidirectional transmissions is the Sensaphonics 3D Ambience In Ear Monitor design (see: http://www.sensaphonics.com/monitors.html).

These in ear monitors let each musician hear his/her own performance and by providing 26db of sound isolation and excellent sound quality, reduce the risk of hearing loss. With Sensaphonic monitors, the user will be able to hear the mix at a lower volume, and protect the hearing from prolonged high-decibel sound exposure. An innovative design with built-in quality microphones and electronics on the outside of in-ear monitors, developed by Robert Schulein, allows the user to balance the sound captured in the room, with the sound monitoring the remote room. The total mix level can be adjusted separately by each user. This solution is not inexpensive, especially when there are several participants that need to be equipped with in-ear monitors, but it does provide the best quality and intelligibility of sound with absolutely no echo.

In summary, these results indicate that both echo prevention and suppression can be used and provide improvement in the comfort of communicating speech and music. Stereo and multichannel methods are necessary to improve the quality of stereo and multichannel communication. To achieve best results in echo reduction, and as few negative side effects as possible, one should combine acoustic methods of echo avoidance with electronic methods of echo suppression, or cancellation, to achieve the most effective removal of echo once it has been transmitted.

We will keep observing the emerging developments in multichannel echo cancellation and test them once they become available for use in advanced communication systems.

Figures and tables below:

Table 1: Subjective evaluation, perceived signal quality (of partner voice)

	USER RATING (1= least, 5= most)			
SYSTEM/STATE	User 1	User 2	User 3	User 4
Gentner				
Bypassed	4	3	3	3
Engaged	3	3	2	4
MIO				
Bypassed	3	4	4	4
Engaged	3	3	1	2
Bronto				
Bypassed	4	3	4	4
Engaged	1.5	1	2	2

Table 2 : Subjective evaluation, perceived echo audibility (of own voice)

	USER RATING (1= least, 5= most)			
SYSTEM/STATE	User 1	User 2	User 3	User 4
Gentner				
Bypassed	5	2	3	4
Engaged	3	1	3	4
MIO				
Bypassed	5	4	2	3
Engaged	1	1	2	1
Bronto				
Bypassed	4.5	4.5	3	5
Engaged	4	3	3.5	3

Table 3 : RASTI values <u>without 300Hz tone</u> (STI)

	Bypassed		Engaged
GENTNER[2]		**GENTNER**[2]	
HatsL	0.82	HatsL	0.83
HatsR	0.81	HatsR	0.87
MIO[1]		**MIO**[1]	
HatsL	0.64	HatsL	0.90
HatsR	0.67	HatsR	0.90
BRONTO[3]		**BRONTO**[3]	
HatsL	0.73	HatsL	0.87
HatsR	0.73	HatsR	0.85

[1]Measured June 15th [2]Measured June 20th [3]Measured June 28h

Table 4 : RASTI values <u>with 300Hz tone</u> (STI)

	Bypassed		Engaged
GENTNER[2]		**GENTNER**[2]	
HatsL	0.84	HatsL	0.89
HatsR	0.86	HatsR	0.89
MIO[2]		**MIO**[2]	
HatsL	0.82	HatsL	0.79
HatsR	0.84	HatsR	0.89
BRONTO[3]		**BRONTO**[3]	
HatsL	0.73	HatsL	0.87
HatsR	0.74	HatsR	0.86

[1]Measured June 15th [2]Measured June 20th [3]Measured June 28h

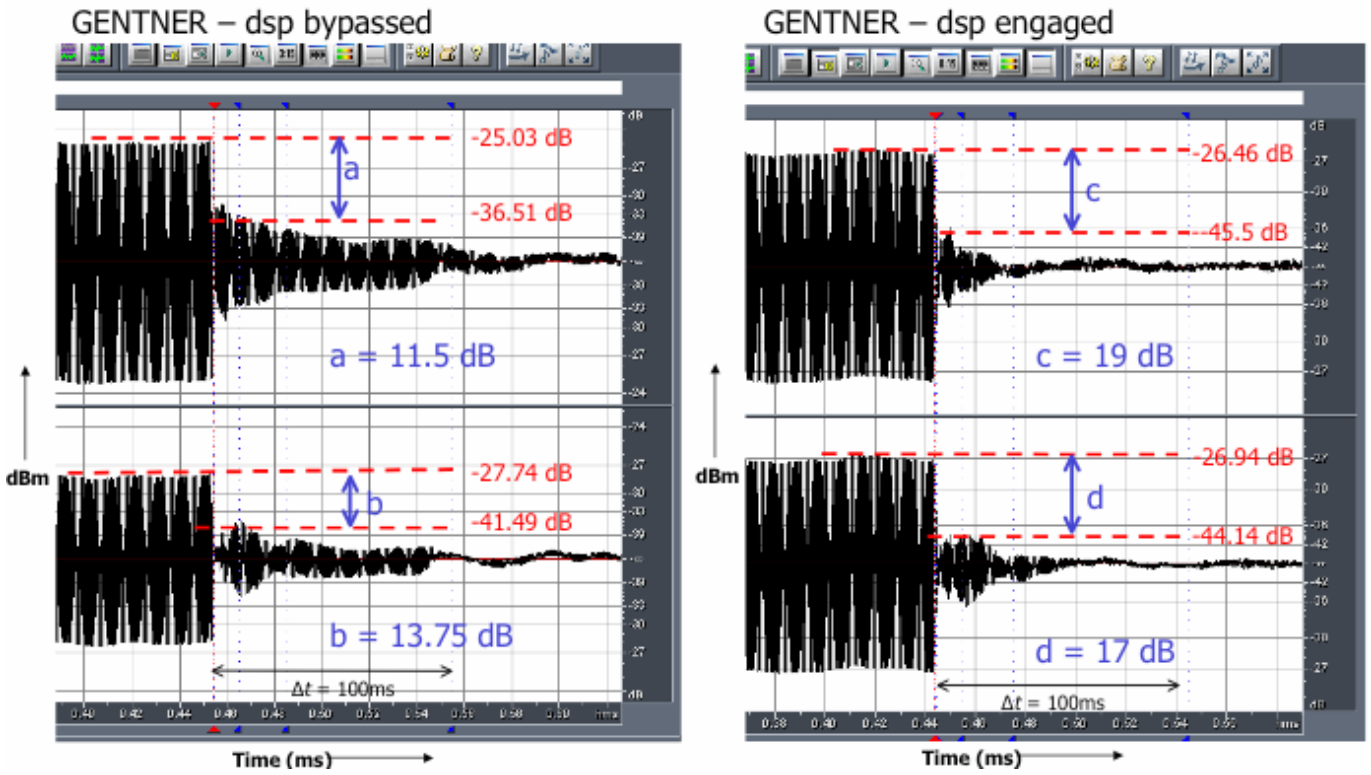

Figure 2: Gentner system, showing 7.5, 3.75 dB (L-R, maximum RMS over 100ms sample) difference between bypassed and echo suppression modes.

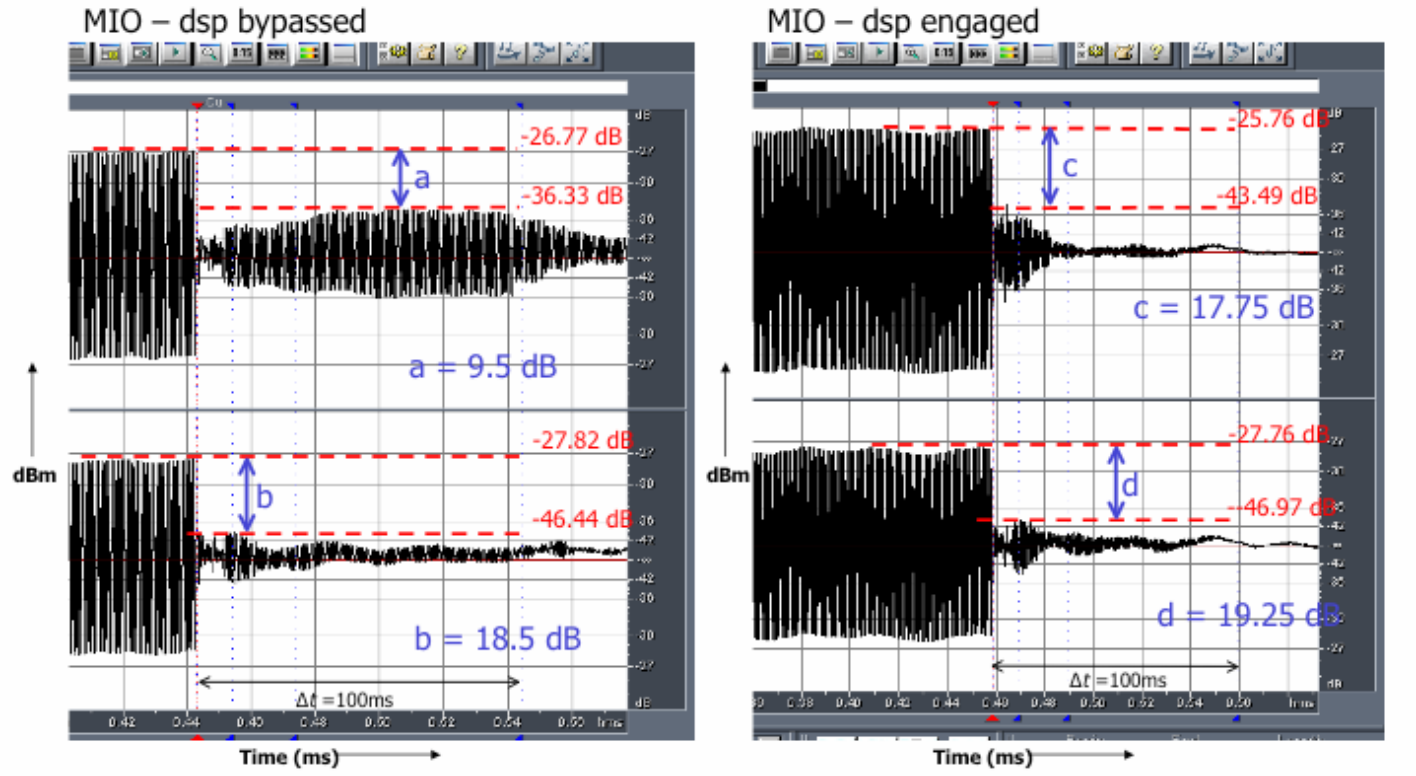

Figure 3: Metric Halo system, showing 8.25, 0.75 dB (L-R, maximum RMS over 100ms sample) difference between bypassed and echo suppression modes.

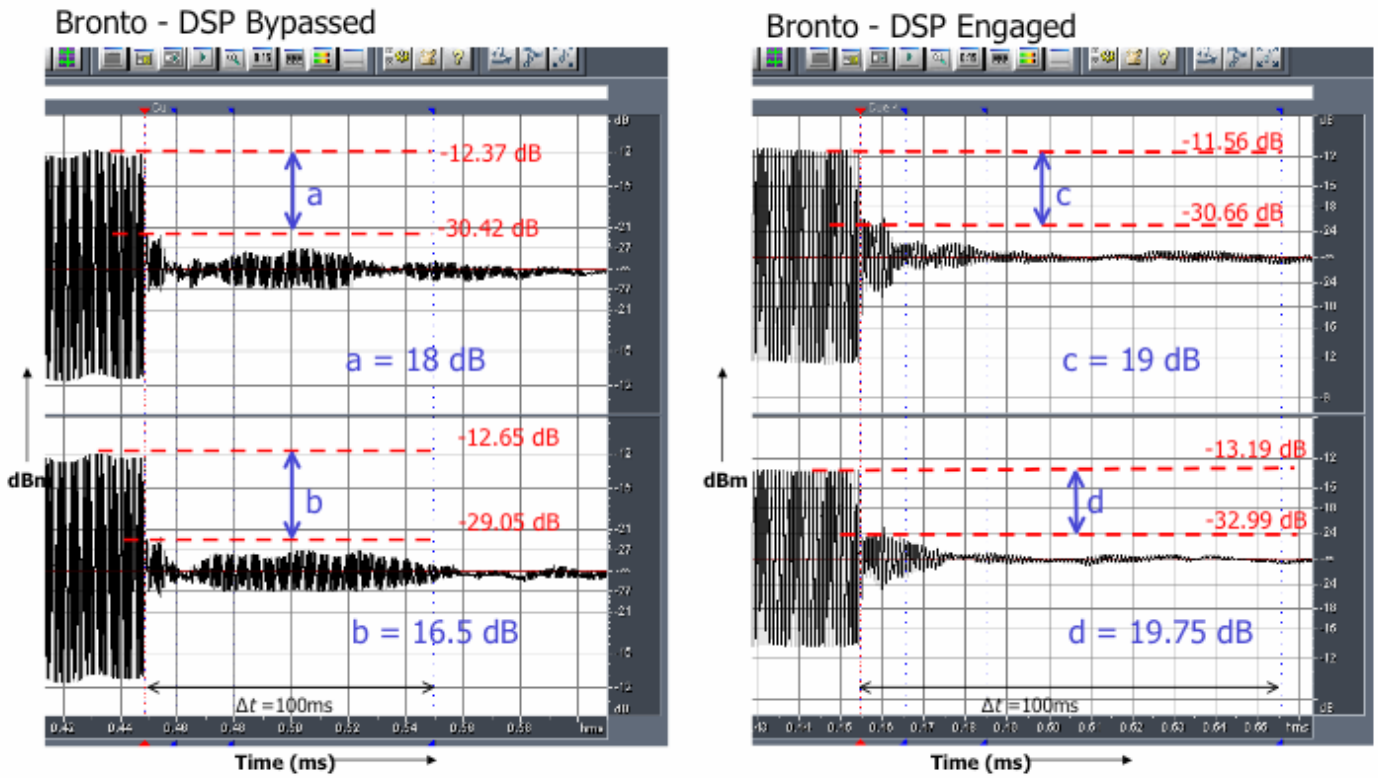

Figure 4: Bronto system, showing 1, 3.25 dB (L-R, maximum RMS over 100ms sample) difference between bypassed and echo suppression modes.

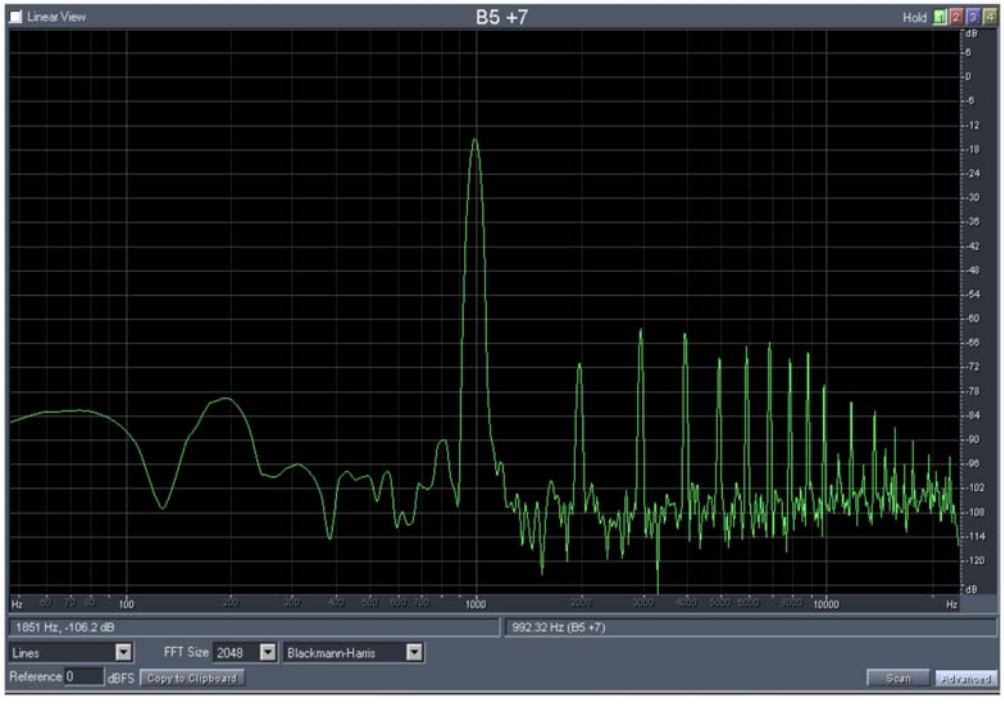

Figure 5: FFT plot of sine tone used in ping tests .

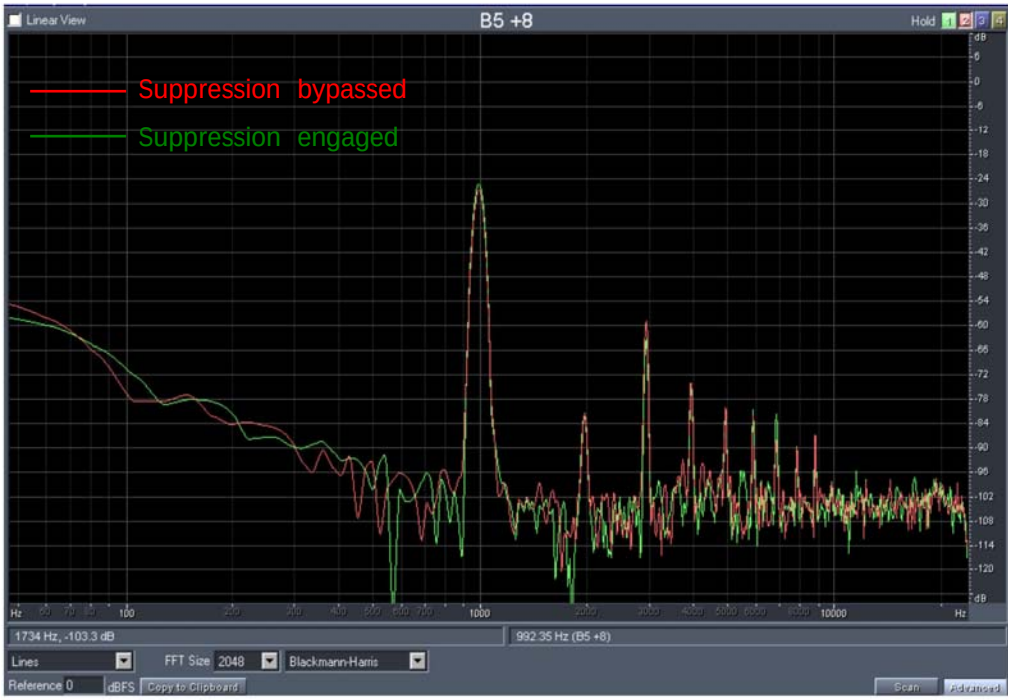

Figure 6: FFT plot of Gentner System, HATS left, ping tone plus decay envelope.

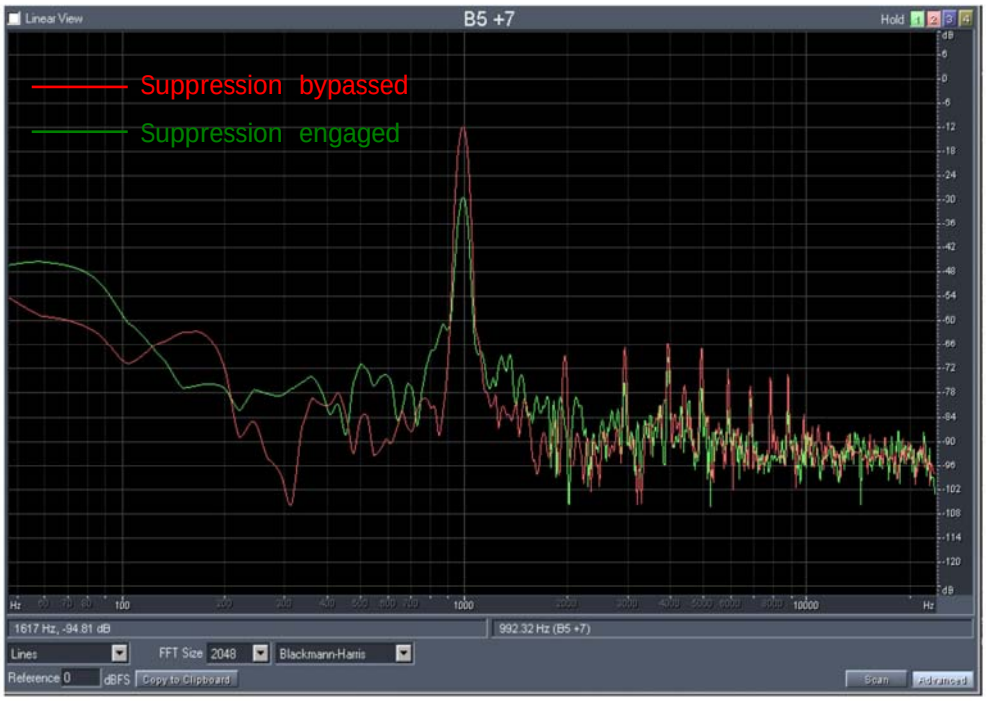

Figure 7: FFT plot of Gentner System, HATS left, decay envelope only.

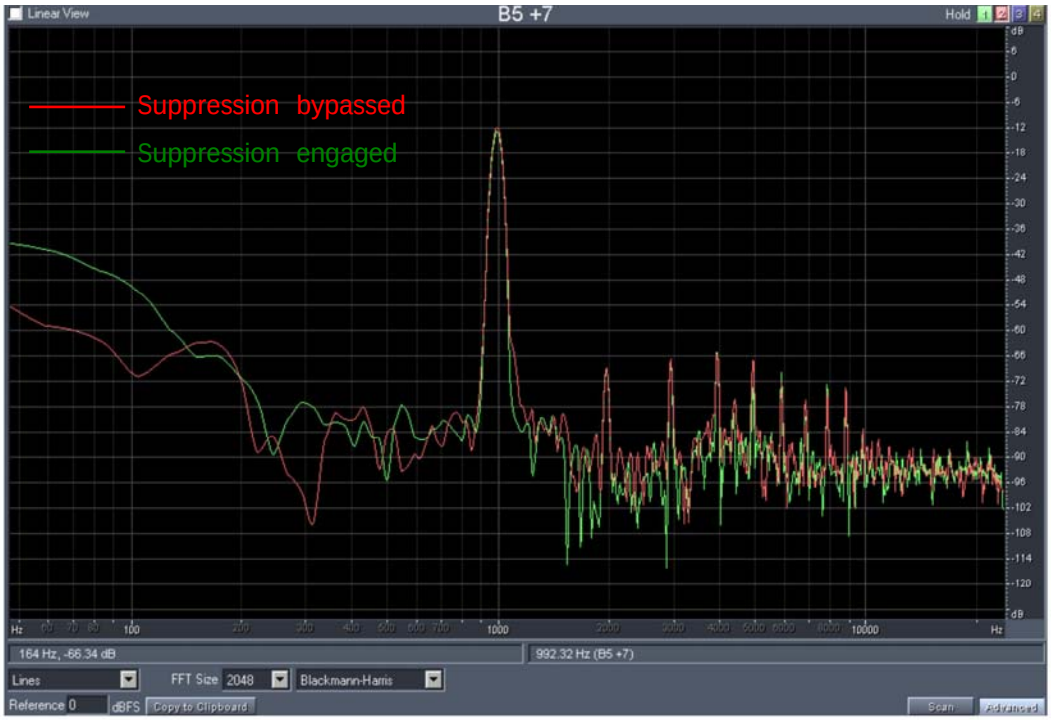

Figure 8: FFT plot of MIO System, HATS left, ping tone plus decay envelope

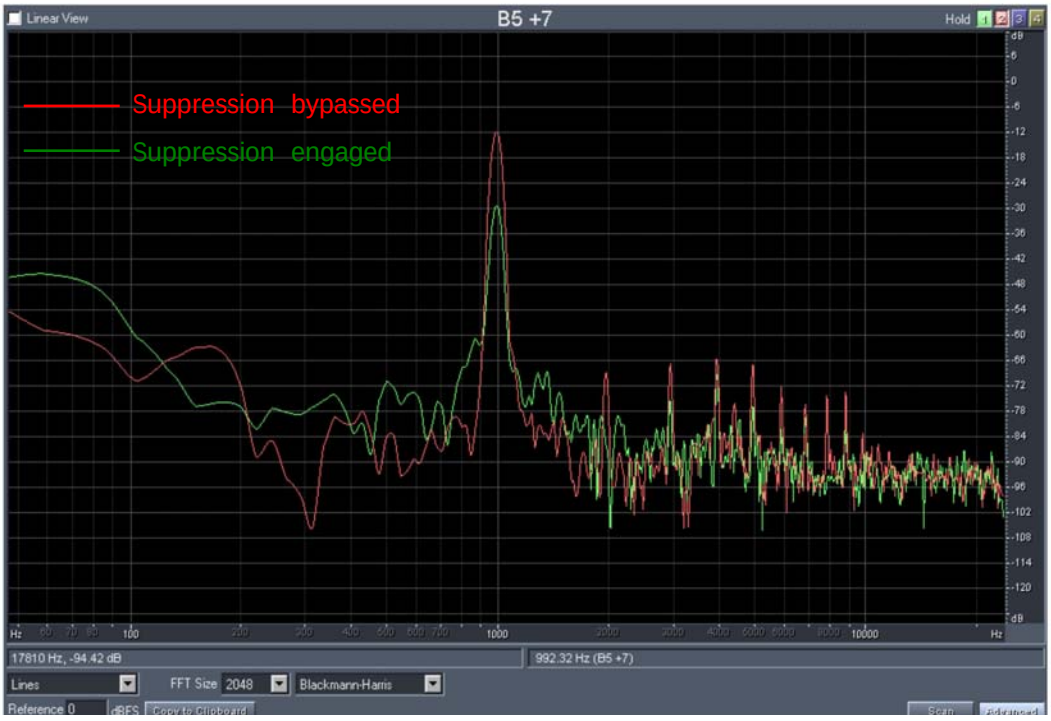

Figure 9: FFT plot of MIO System, HATS left, decay envelope only

Figure 10: FFT plot of Bronto System, HATS left, ping tone plus decay envelope

Figure 11: FFT plot of Bronto System, HATS left, decay envelope only

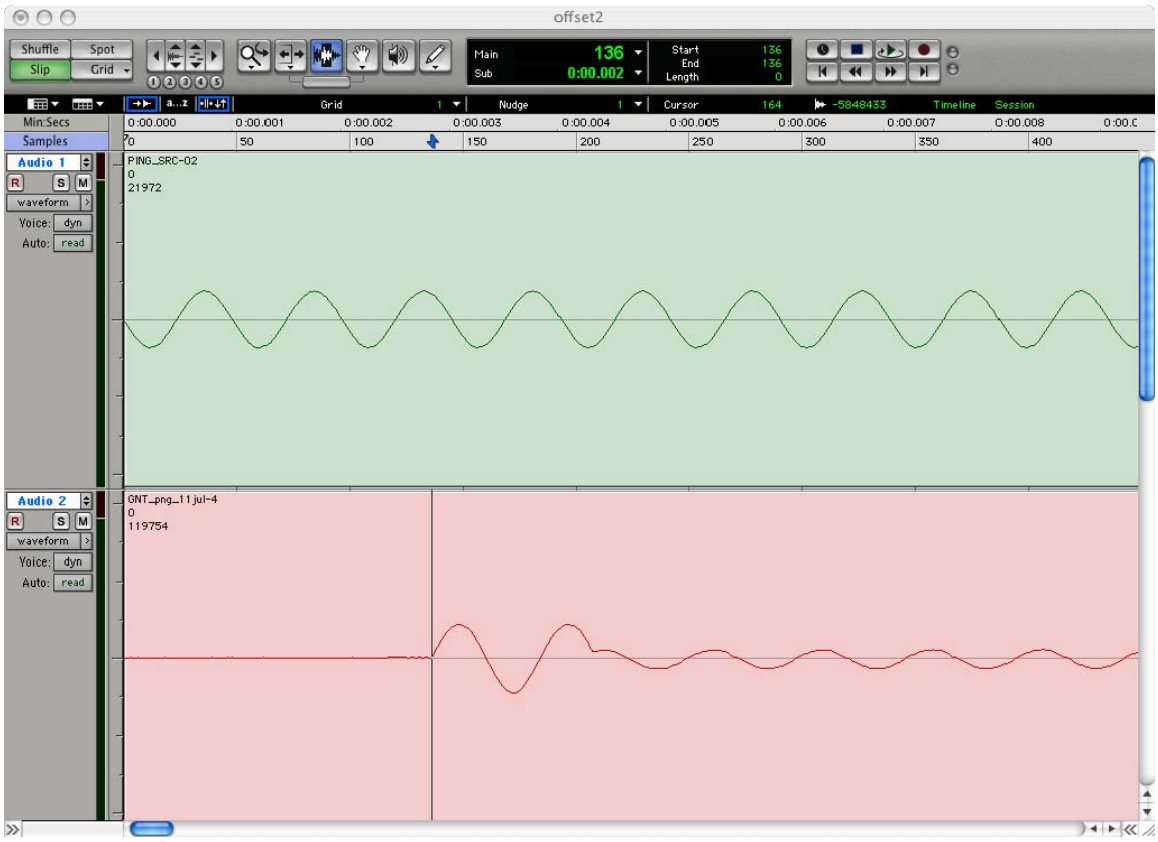

Figure 12: Hardware Ping, Gentner w/o Suppression. 136 samples offset (net offset minus MIO a/d processing = 74 samples latency)

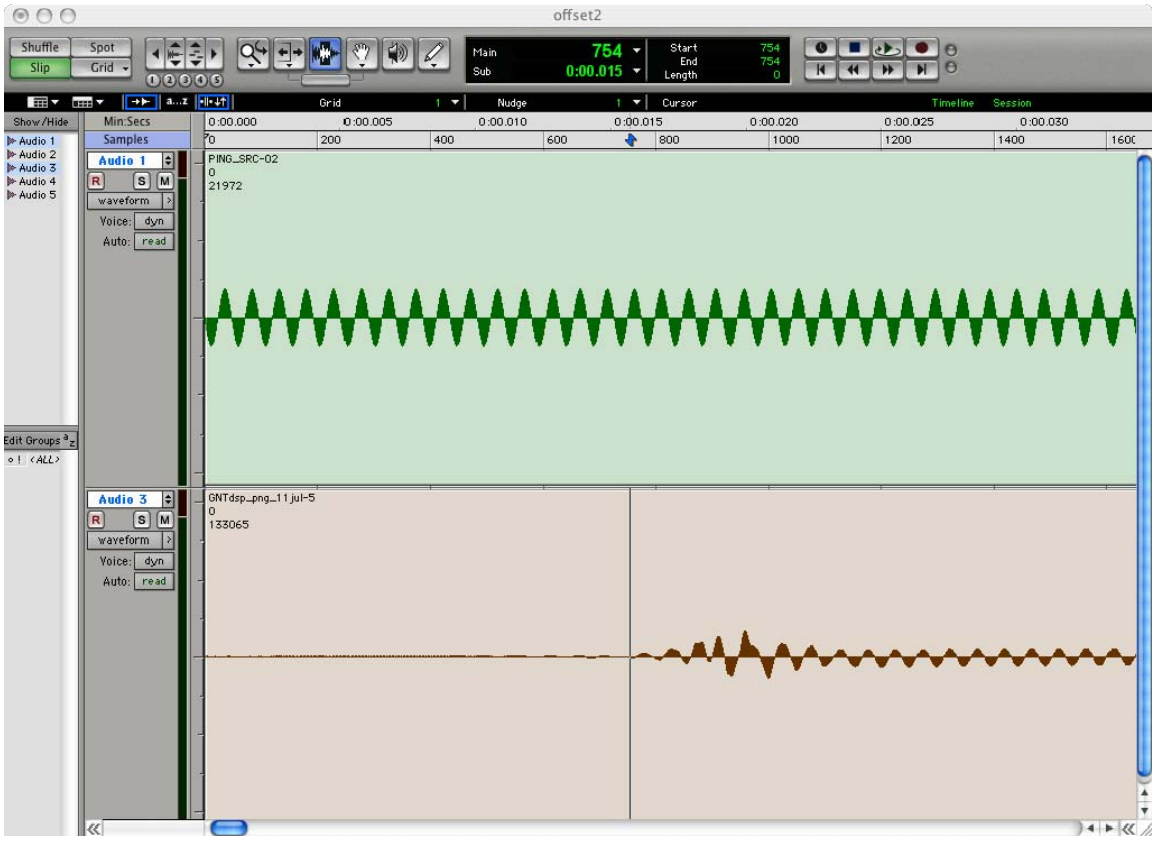

Figure 13: Hardware Ping, Gentner w/suppression. 754 measured offset (net offset minus MIO a/d processing = 690 samples latency)

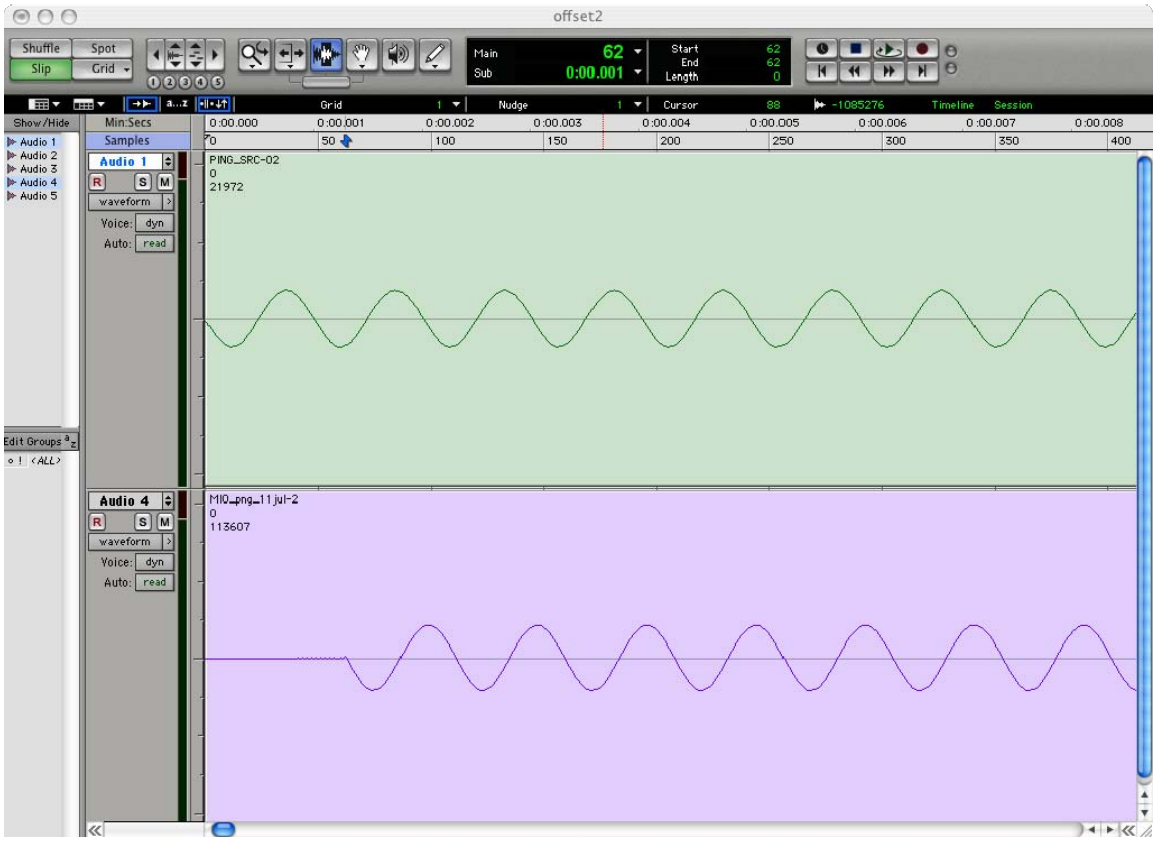

Figure 14: Hardware Ping, MIO w/o suppression, 62 samples latency.

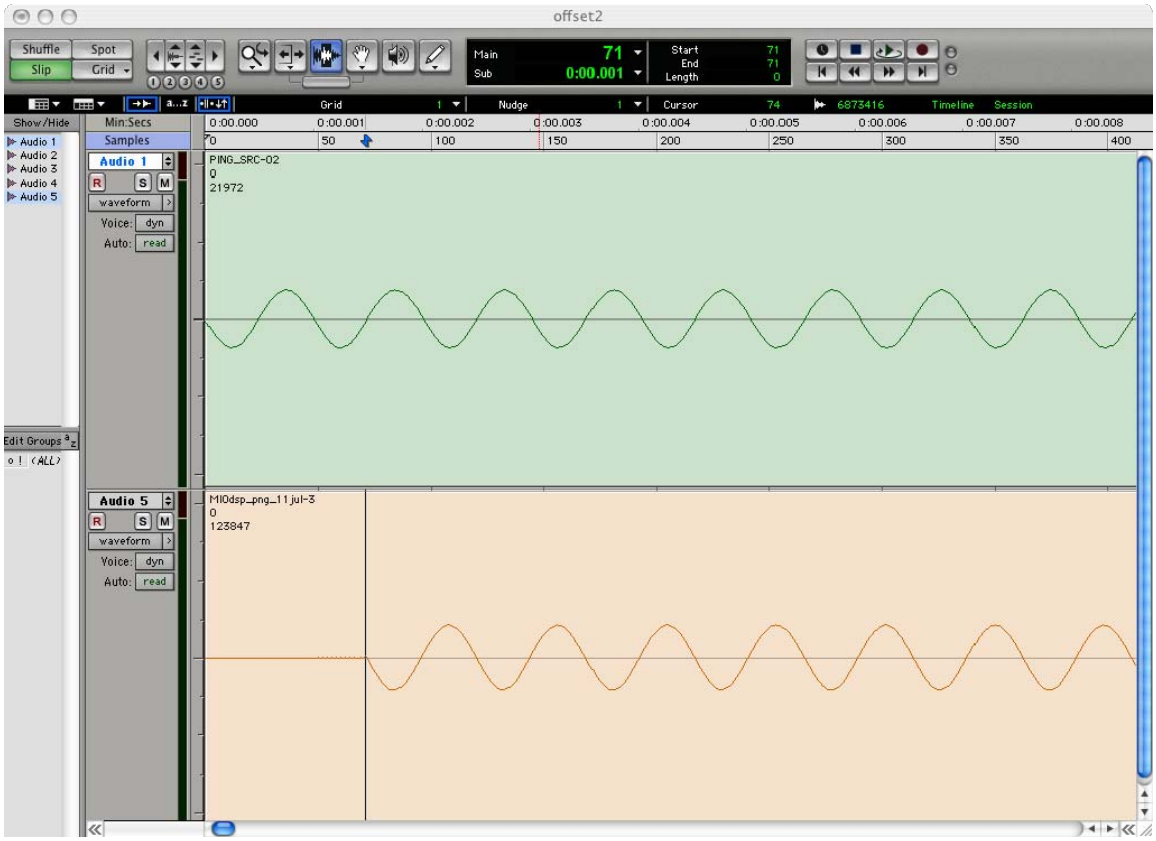

Figure 15: Hardware Ping, MIO w/suppression, 71 samples latency.

Figure 16: Immersive Presence Laboratory (IPL) – general view of loudspeakers and of the listening area.

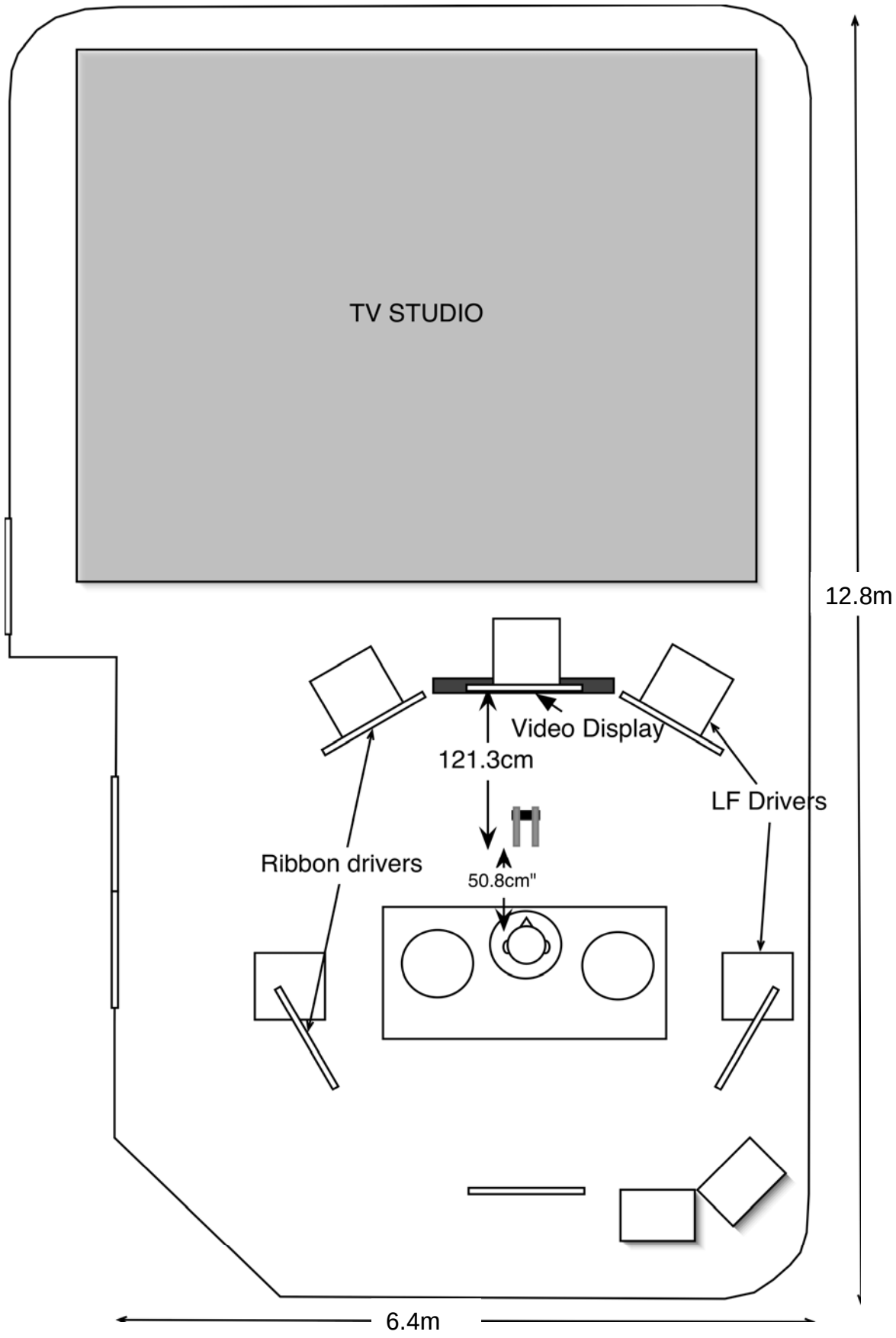

Figure 17: Overhead view of IPL with subject position and relative distances (NTS)

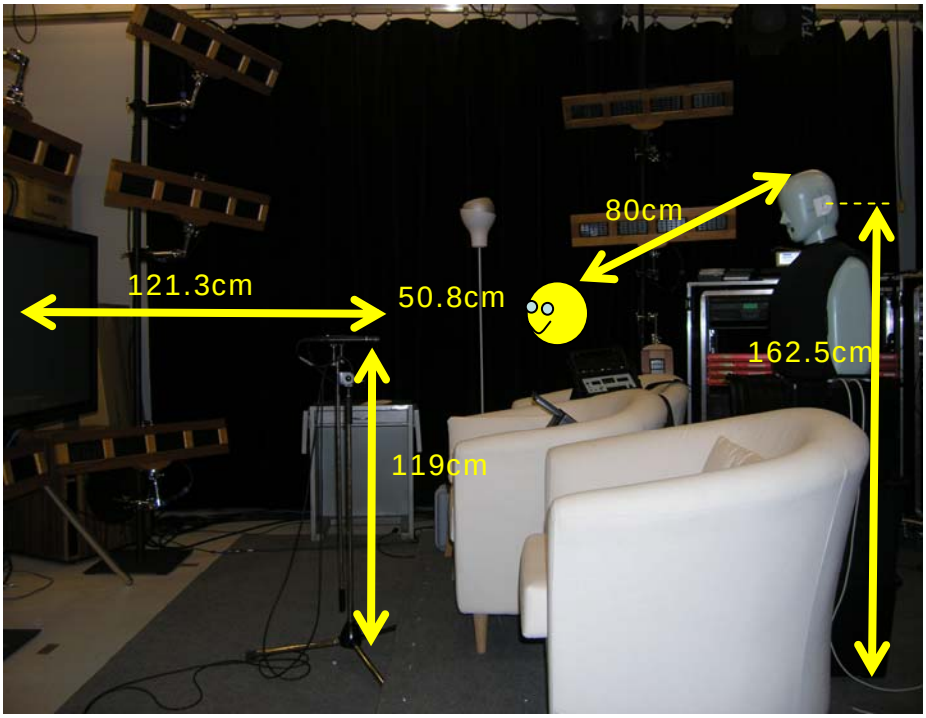

Figure 18: Setup in IPL showing speaking subject's and artificial listener's locations, and distanced to the microphones and the loudspeakers.

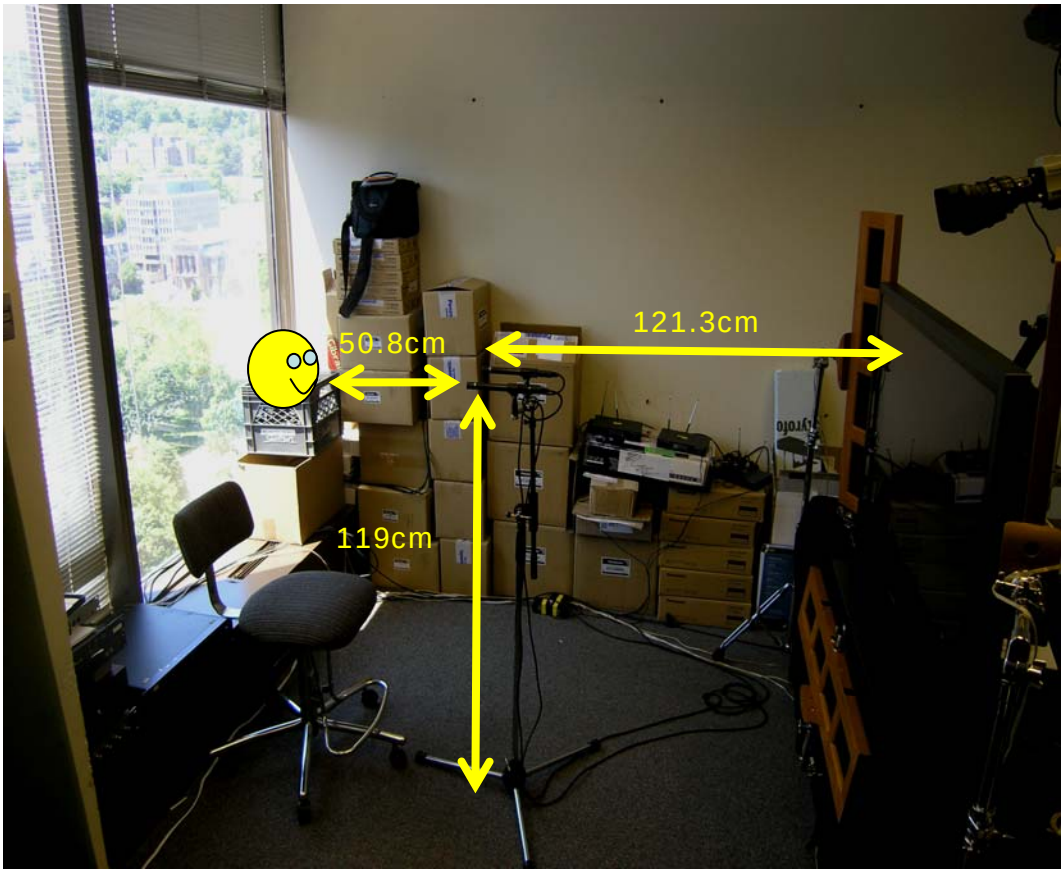

Figure 19: Set up in room 1684 with subject's position and relative distances to the microphones and the loudspeakers.

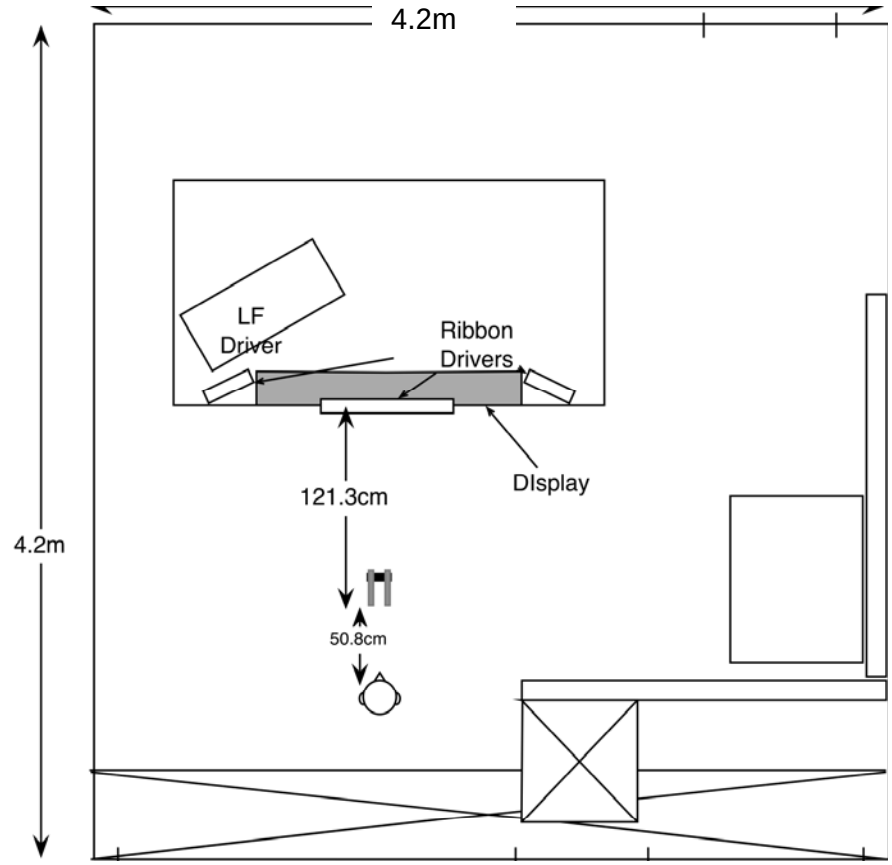

Figure 20: Overhead view of room 1684 with subject position and relative distances (NTS)

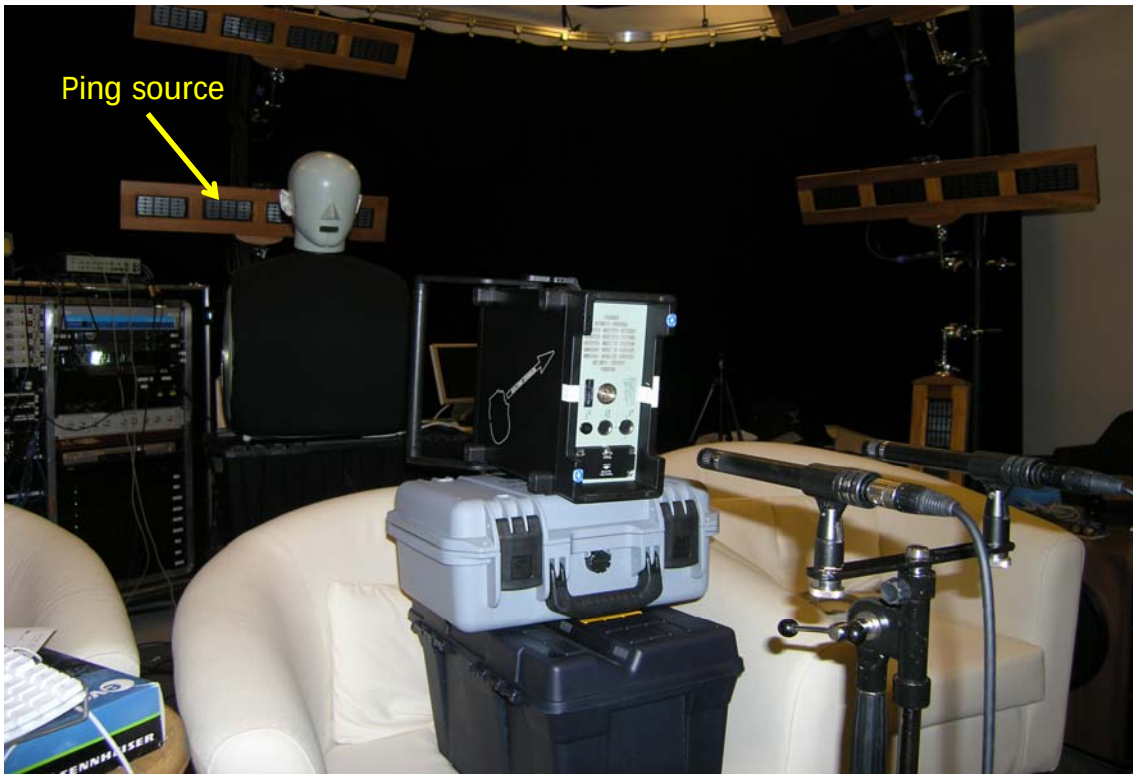

Figure 21: Artificial sound source simulating a speaking person (part of the B&K RASTI system) set up in front of the two microphones. Behind, the artificial listener and the loudspeaker sending the ping tone.

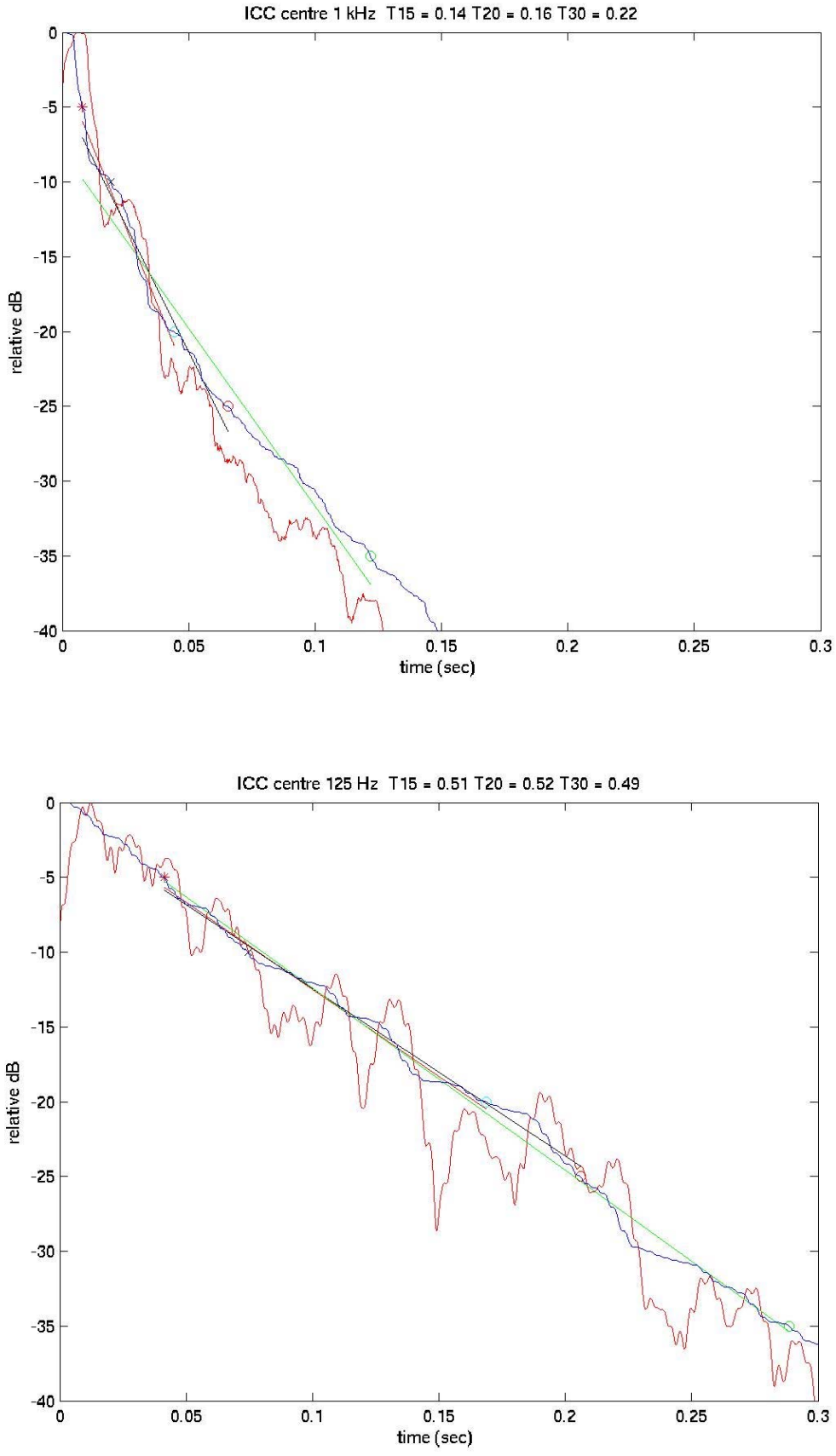

Figure 22: Reverberation decay curves of the IPL (part of TV studio) at 1kHz (upper graph) and at 125Hz (lower graph) showing greater absorption of high frequencies by the velour curtains.

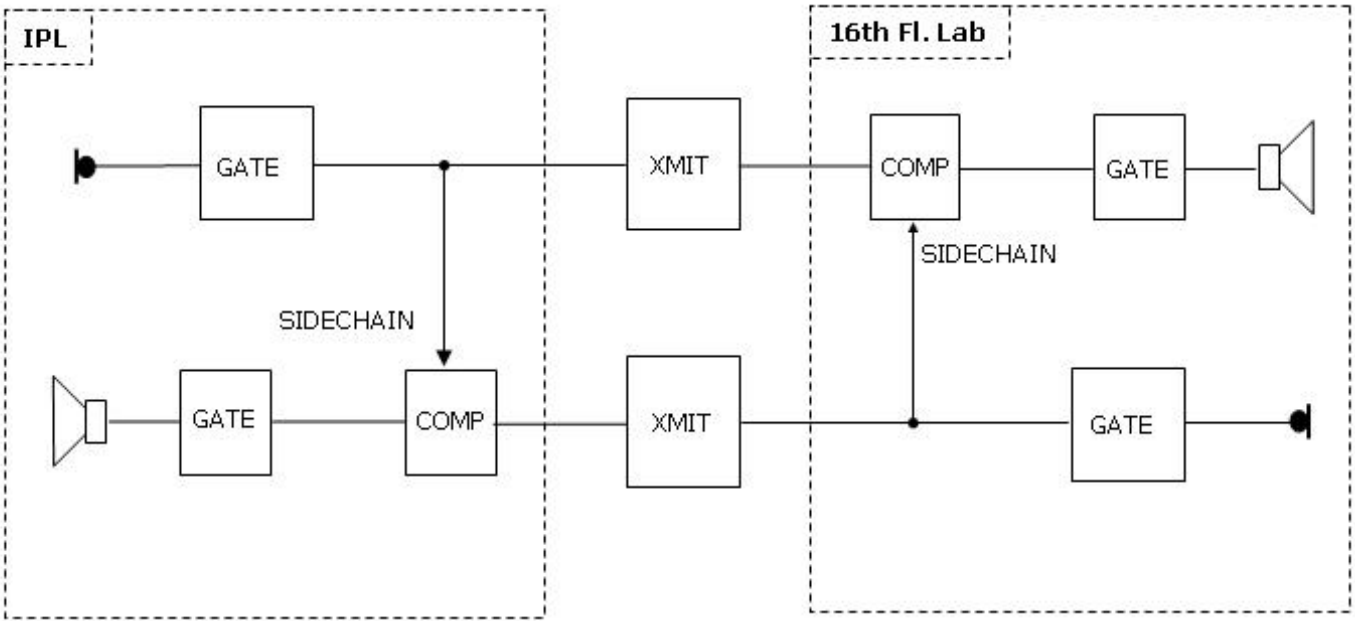

Figure 23: Metric Halo Compression/Gating scheme developed by Woszczyk and employed in this experiment to reduce the formation of echo in bidirectional transmission. Any number of the Comp/Gate processing cells (left or right dotted outline) can be set up for any number of Mic/LS pairs, and be cross-linked via the side-chain inputs. Upon sensing the presence of a microphone signal, the compressor lowers the gain of the microphone signal of the opposite (returning) side and thus decreases the gain in its feedback loop. This automatic gain riding only affects the gain of the microphone of the non-speaking person. The assumption is that only one person speaks at a time. When two sides speak simultaneously, the loop gain is reduced for both sides. This dynamic aspect of gain riding may affect their ability to hear each other. The dynamic action of each Comp/Gate cell has to be properly adjusted for speech and music to reduce the sometimes audible pumping effect (audible change of background level). A quiet room with low reverberation time is a desired at each side, and/or close microphone placement to the source, to assure the best results. The gate on the microphone side helps to lower the level of room noise and reverberation, and thus the audibility of the pumping effect.

References

1. The Recording Studio that Spanned a Continent. Cooperstock, J.R. and Spackman, S. (2001) IEEE International Conference on Web Delivering of Music (WEDELMUSIC), Florence.

2. Real-Time Streaming of Multichannel Audio Data over Internet. Xu, A., Woszczyk, W., Settel, Z., Pennycook, B., Rowe, R., Galanter, P., Bary, J., Martin, G., Corey, J., and Cooperstock, J.R. (2000) Journal of the Audio Engineering Society, July-August.

3. A multi-filter approach to acoustic echo cancellation for teleconferencing. Usher, J., Cooperstock, J.R., and Woszczyk, W. (2004) 75th Meeting of the Acoustical Society of America, New York, May 24-28.