Technical Support Guide

HARMAN
A SAMSUNG COMPANY

v1.0

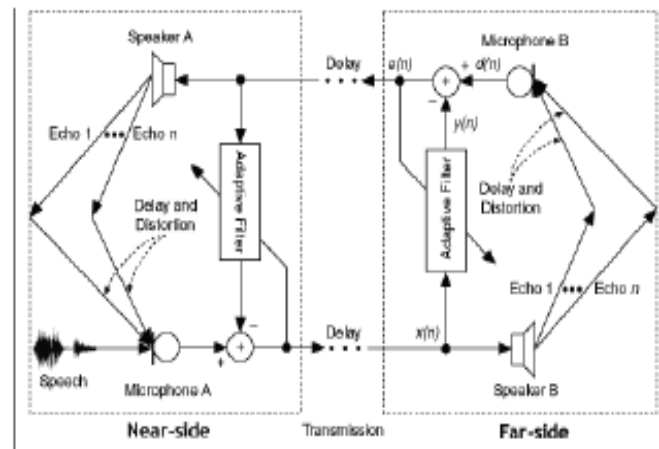ACOUSTIC ECHO CANCELLATION

22-Apr-20

What is acoustic echo and why does it need to be cancelled?

Acoustic echo occurs in a conferencing system when the far-side speech played in local loudspeakers is picked up by microphones in the near-side room and is transmitted back to the far side. This transmitted signal is a delayed version of the original, which causes the echo.

The received far-side signal does not transfer directly from the speaker to the microphone, but is subject to the artifacts of the room. This may include differing signal paths causing reverb, frequency filtering and attenuation. These effects are the result of the transfer function of the room. This transfer function is also dynamic as objects in the room move or the microphone moves position.

To subtract the unwanted signal correctly, the Acoustic Echo Cancellation (AEC) processor needs to simulate the dynamic room transfer function. It can then apply that transfer function to the received signal and correctly subtract the modified original signal.

Each Soundweb London AEC input card consists of four AEC input channels.

Each channel offers the following features:

- Independent 20Hz - 8kHz algorithm

- Individual AEC references

- Automatic Gain Control

- Noise Cancellation

- Adaptive (Speech Passing) Non-Linear Processing

- Extremely fast convergence rates of 49dB/s

AEC Input cards may only be used in the following BSS Audio Soundweb London devices configured for 48kHz operation:

- BLU-100
- BLU-101
- BLU-102
- BLU-120
- BLU-160
- BLU-320
- BLU-325
- BLU-326
- BLU-800
- BLU-805
- BLU-806

AEC DEFINITIONS

Here are a few Acoustic Echo Cancellation terms.

Convergence Rate - A measure of how fast the AEC algorithm can recognize and remove echo from the signal path.

Echo Return Loss (ERL) - A measure of the coupling between the AEC reference signal and the AEC input signal.

Echo Return Loss Enhancement (ERLE) - Shows the loss through the linear AEC algorithm (not including the non-linear processing).

AEC Card Control Panel

The AEC default control panel is ordered in two groups of controls for every input channel. The first group of controls is identical to standard Soundweb London input cards and functions in the same manner. These controls are the audio input meter, configurable as Pre-or Post-AEC, input meter controls, Attack, Release, Reference, and Phantom Power for each input channel. The second group of controls is the AEC controls.

AEC Control Panel [Basic]

The basic AEC control panel allows enabling and disabling of AEC and Automatic Gain Control (AGC), and allows setting levels for noise cancellation and nonlinear processing.
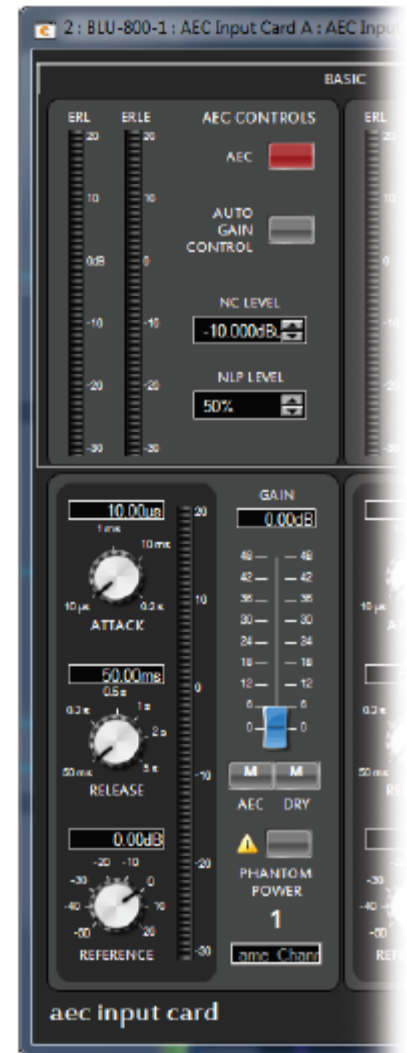
AEC

This button enables or disables AEC processing for each channel. When this button is enabled the AEC algorithm will remove the acoustic echo from the audio channel with linear processing and with a specified amount of non-linear processing. (See Non-Linear Processing Level)

BSS Audio Soundweb London AEC Panel

ERL Meter

The Echo Return Loss (ERL) meter is a measure of the room's natural attenuation of the far-side audio as it leaves the loudspeaker(s) and re-enters the microphone(s). This parameter is controlled by proper gain structure setup, ensuring a good signal to noise ratio and reasonable headroom for the AEC input signal. A proper gain structure is critical for distortion free sound and optimal performance for AEC. This is the single most important parameter when setting up the AEC system.

The AEC algorithm will best recognize and remove the acoustic echo when this meter is displaying in the green range, which is indicated on the control panel below 10dB. The algorithm will continue to converge over 10dB, but the convergence rate will decrease in that range. This meter will not update during double-talk where both far-side and near-side signals are present. It is updated based on far-side speech only.

ERLE Meter

The Echo Return Loss Enhancement (ERLE) Meter measures how much acoustic echo is being removed from the signal path. This measurement consists of the natural room attenuation as indicated by the Echo Return Loss (ERL) meter and the amount of echo removed by the AEC algorithm. A lower signal indicates more echo being removed.

NOTE: As dictated by industry standards, Non-Linear Processing contributions are not included in this guide. Non-Linear Processing contributions are made in addition to this meter's reading.

Non-Linear Processing Level

The Non-Linear Processing (NLP) setting determines the amount of non-linear suppression that will be applied in conjunction with the AEC algorithm for each channel. Non-Linear Processing will remove the residual echo not removed by the linear part of the AEC algorithm. This parameter represents a trade-off between achieving good double-talk performance, with no suppression of the local speech signal, and very robust echo suppression, with no echo audible on the far side. At its most aggressive setting (NLP at 100%), the non-linear processing will remove any of the residual far-side echo picked up by the microphone. However, this is done with an increased risk that some of the near-side speech will be degraded as well, especially during double-talk. At its least aggressive setting (NLP at 0%), the non-linear processing is effectively disabled, which may let some echo through, but will allow for a more natural double-talk performance. The best setting for this parameter depends upon several issues including the acoustic properties of the room and user preference. The default value of 50% provides a good balance between these two competing factors.

Noise Cancellation Level

The Noise Cancellation (NC) setting will determine the amount of noise cancellation that will be applied to each channel. The noise cancellation algorithm is a very advanced algorithm that will remove steady-state noise without compromising the quality of speech passing through the channel. It is very effective for removing projector noise, HVAC, and other unwanted background noise that compromises speech intelligibility.

AEC Control Panel [Advanced]

The advanced panel gives access to controls for the Automatic Gain Control and Signal Threshold features.

Automatic Gain Control

The Automatic Gain Control (AGC) is designed for voice applications. It compensates for varying speech levels at the near end only, providing the far end with a more consistent signal level.

To use the Automatic Gain Control, first define target levels for the transmitted speech signal. The default target levels for AGC are a maximum of 6dBu and a minimum of -10dBu. This provides a target window with 16dB of dynamic range. If the speech level is within the target window already, the AGC-applied gain will be set to 0 dB.

If the speech signal drops below the minimum target level, the AGC will increase the gain to compensate. On the upper end, the AGC will limit the gain it can add to a signal by a maximum gain setting. Once the AGC has adjusted its gain to meet the maximum gain setting, it will stop adding gain, even if the minimum target level is not reached.

Setting the maximum gain too high can cause inconsistent gain structures and bring up the noise floor. Similarly, if the speech signal level is higher than the maximum target level, the AGC will lower the speech level down to the maximum target level. A generous range for the maximum and minimum gain has been provided. Care should be taken, particularly with the maximum gain setting, to avoid extreme levels. Situations where the maximum gain setting would be set over 10dB will be rare.

Attack and Release rates for the AGC describe how fast the gain will be adjusted, and the AGC meter shows the current amount of gain being applied to the signal.

Signal Threshold

In a conferencing system, some microphones have a mute or push-to-talk feature. When the mute status of a microphone changes, the characteristics of the conferencing system change and echo may leak through as the AEC reconverges. A signal threshold is defined to allow the mute or push-to-talk feature to work seamlessly with AEC. Using the threshold, a level is defined that is below the normal, ambient noise floor of the room. If the microphone level goes below this level, the AEC algorithm will treat the microphone as muted, and minimize any echo that would have occurred otherwise. The "Active" LED indicates that the microphone level is over the threshold, and the microphone is not muted. When the LED is off, the microphone level is below the threshold, and the microphone will be treated as muted.


BSS Audio Soundweb London AGC Panel

To set the Threshold:

- Mute the microphone.

- Set the Threshold to a level where the Active LED turns on and off intermittently.

- Raise the threshold above this level by 3 to 6 dB. The LED should now be off. Un-mute the microphone, the LED should light. This process may need to be repeated if the microphone's preamp gain is adjusted. To disable the mute feature based on signal threshold, set the Threshold to its minimum value.

HiQnet Audio Architect AEC card configuration

Each recognized AEC card in a Soundweb London unit will appear in the Venue view as shown here:

The left hand 'AEC Input Card' block functions like a standard Soundweb London input card. This block contains the four channels of processed AEC audio as well four channels of 'dry' or unprocessed input audio being fed into the Soundweb London AEC card.
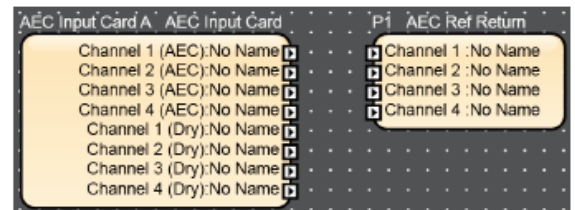
The right hand block 'AEC Input Ref Return' is used to provide the reference signal for each AEC algorithm. This is the signal that will be removed by the AEC algorithm from the signal path, and should be taken from as close to the output as possible. This will provide the AEC algorithm with the most accurate representation of the signal to be cancelled, and provide the best AEC performance.
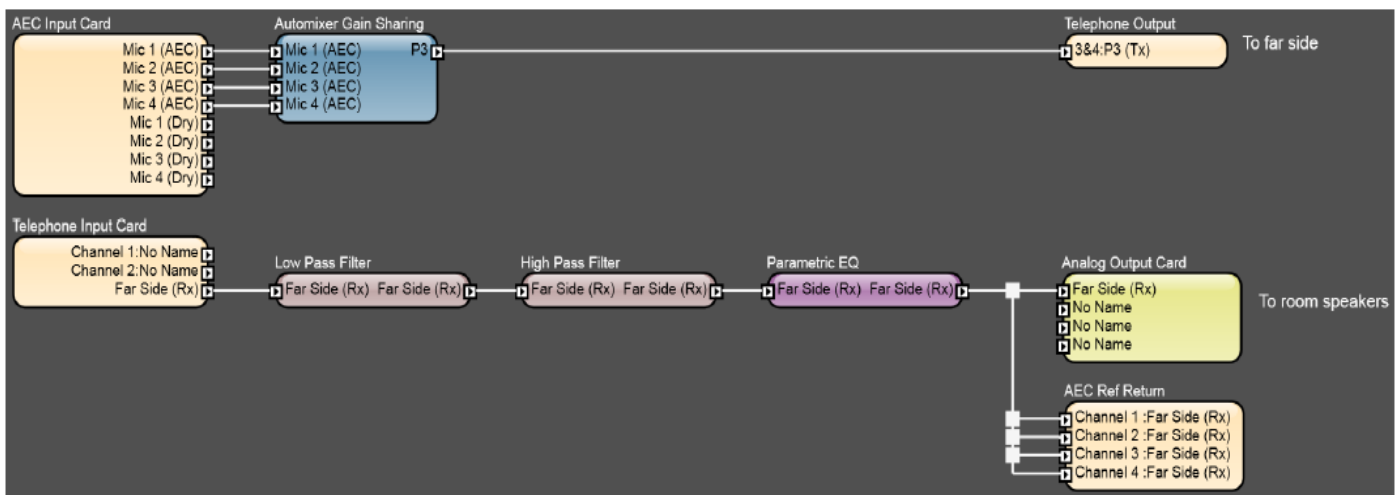
EXAMPLE: BASIC CONFERENCING WITHOUT LOCAL SOUND REINFORCEMENT

Local Sound Reinforcement refers to a design where the local microphones feed both the far-side and near-side speakers. This typically applies in large rooms where other participants located in the same room are unable to hear the person speaking.
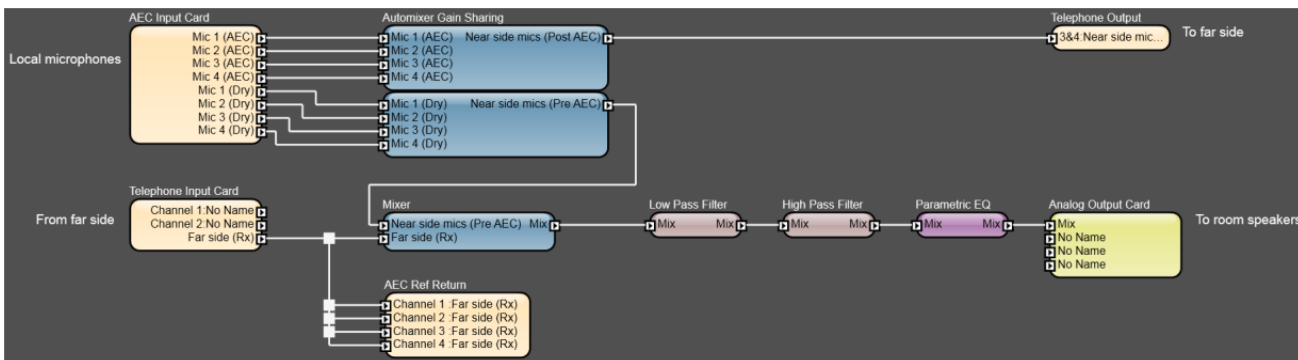
This example shows four microphones feeding the local audio to the far side via a Telephone Hybrid card. The far-side audio is received via the Telephone Hybrid and processed by Soundweb London Low Pass, High Pass and Parametric EQ processors before being sent to the local room's speaker(s). Once the far-side audio leaves the local speakers, the signal will bounce around the room and re-enter the local microphone, mixing with the local side's speech. The far-side signal will be sent back to the far side resulting in an unwanted echo.

To prevent this, the far-side signal is sent to the Reference inputs of the AEC card where the AEC algorithm will compare this signal with the microphone input signals of the AEC card and remove the far-side reference signal. This eliminates the unwanted echo because only the local side audio is being sent to the far side.

In order for this design to function properly the AEC Reference Return inputs must be taken after the room processing blocks. If the Reference is taken prior to the processing, the AEC algorithm will not understand that certain frequencies were cut / boosted intentionally and will not be able to model the room to its full ability.

# EXAMPLE: BASIC CONFERENCING WITH LOCAL SOUND REINFORCEMENT

This example shows four microphones feeding the audio to the far side via a Telephone Hybrid card as well as feeding the local speakers for local sound reinforcement. Signal mixing is performed using the Gain Sharing Automixer processing object. The best method for this type of design incorporates a mix-minus setup to maintain proper gain structure, and to prevent the speaker directly above the person talking from transmitting a room-colored copy that will re-enter the open microphone and be transmitted to the far side along with the original voice signal.

The design below shows both the far-side and near-side signals feeding the local room speakers. This design works, but as explained previously, the AEC algorithm will not perform to its full potential.

If the Reference is moved to the same location as in the previous 'No Local Sound Reinforcement' example, it will satisfy the rule of placing the Reference as close as possible to the speaker output, but in doing so the Reference will be fed with a mix of both the near-side and far-side signals.

Since the Reference signal is the signal to be removed from the input audio path, the AEC algorithm will cancel the microphone signal at the AEC Input Card. Since the input microphone signal path is being fed to the far-side as well as the local loudspeakers, listeners at the far side would not be able to hear the microphone signal. Only a portion of the microphone signal is cancelled because of the Voice Activity Detection state. The Voice Activity Detection processing determines whether the audio is speech or silence / background noise. This would cause the microphone signal to distort both locally and at the far side.
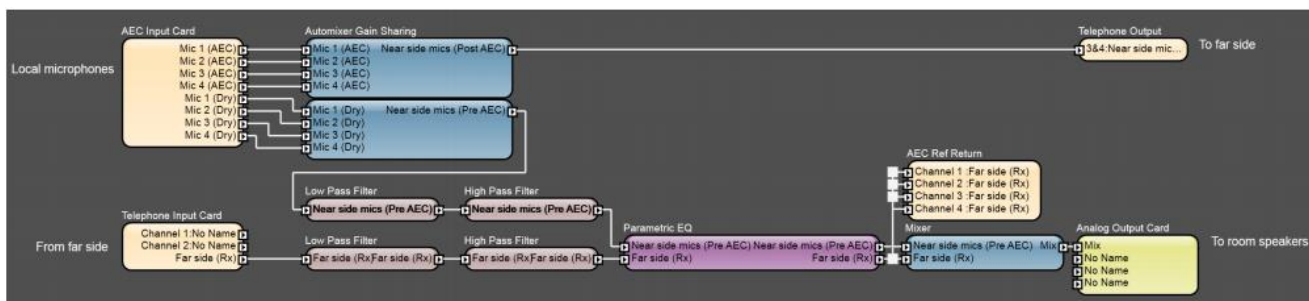
To solve this, another set of High / Low Pass processing objects and a Parametric EQ are utilized, in order to provide the Reference the same signal as the room speaker. This results in only the far-side signal being Referenced and removed while still feeding a mix of both near-side and far-side audio to the room
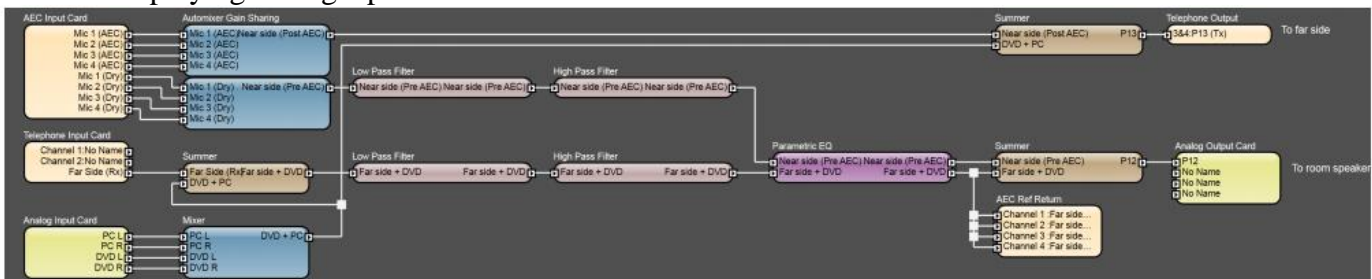
loudspeakers. It is essential that the same settings are maintained in both signal paths. In particular, care must be taken that any non-linear processing such as compression or limiting employed on the speaker output signal, is also applied to the Reference signal. The Copy Parameter Values feature should be employed to ensure that the settings are identical.

EXAMPLE: LOCAL MEDIA DISTRIBUTION

This example adds a DVD player and PC audio outputs to the previous Local Sound Reinforcement design.

The design requires that the local DVD player and PC audio inputs are sent directly to the far side via a Telephone Hybrid card to provide a high quality audio signal.

For the near side to hear the local media through the loudspeakers without unwanted echo and distortion, the Reference signal is removed from the far-side microphone and local media signals. Again, for optimal functionality in removing unwanted echo, the ERL meter should be in the 'green' zone while the local media sources are playing during a presentation.