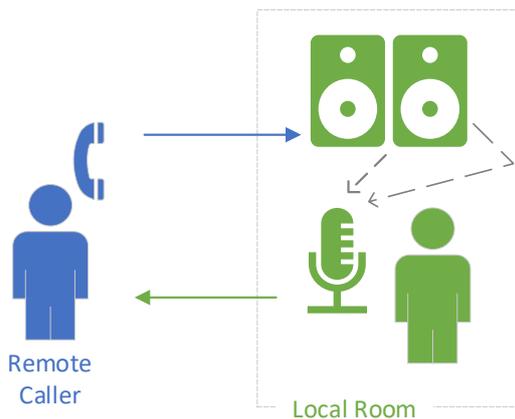


ALLEN & HEATH AEC

ACOUSTIC ECHO CANCELLING MODULE FOR AHM-64

A typical problem in a tele-conferencing or video-conferencing scenario as shown below is the remote caller hearing a delayed copy (echo) of their own speech. This happens because their voice, also known as the 'far-end' signal, is played back via the local room speakers, room impulse response and microphone.

Acoustic Echo Cancellers aim to avoid this by subtracting a simulated echo from the local microphone/s signal/s, also known as the 'near-end'.



AECs can vary in performance and implementation. This document describes a system that has been designed for use in the Allen & Heath AHM audio matrix processors. The functional elements are explained, followed by an overview of the user interface.

M-AHM-64 expansion module

This processing expansion module provides 12 channels of AEC, with each channel having 150ms of echo cancellation. The module utilises two Xilinx FPGA devices incorporating a dual-core ARM Cortex A9 processor each.

The diagram on the next page illustrates the main components of the system.

The **Adaptive Filter**, continually updated, attempts to match the room impulse response, providing the simulated echo.

A few other blocks aid this process, and improve the overall result:

- An **AGC** (Automatic Gain Control) adjusts the far-end gain attempting to maintain a consistent level from the remote caller. The gain is not altered unless there is speech present in the signal.
- A **Talk-State Detector** identifies voice activity and double-talk (both far-end and near-end are speaking). It determines one of four talk states: near-end speaking, far-end speaking, both speaking, neither speaking. This information is passed to several processing blocks. When double-talk is detected for example, the Adaptive Filter does not update, as it cannot distinguish between the far-end echo and the local talker.
- The **NLP** (Non-Linear Processor) implements further processing to attenuate any residual echo not cancelled by the Adaptive Filter, and reduces the primary echo while the Adaptive Filter is converging. It also incorporates Background Noise Reduction.
- An optional **Ducker** uses talk-state information to reduce the level of residual echoes in the case of far-end only activity.

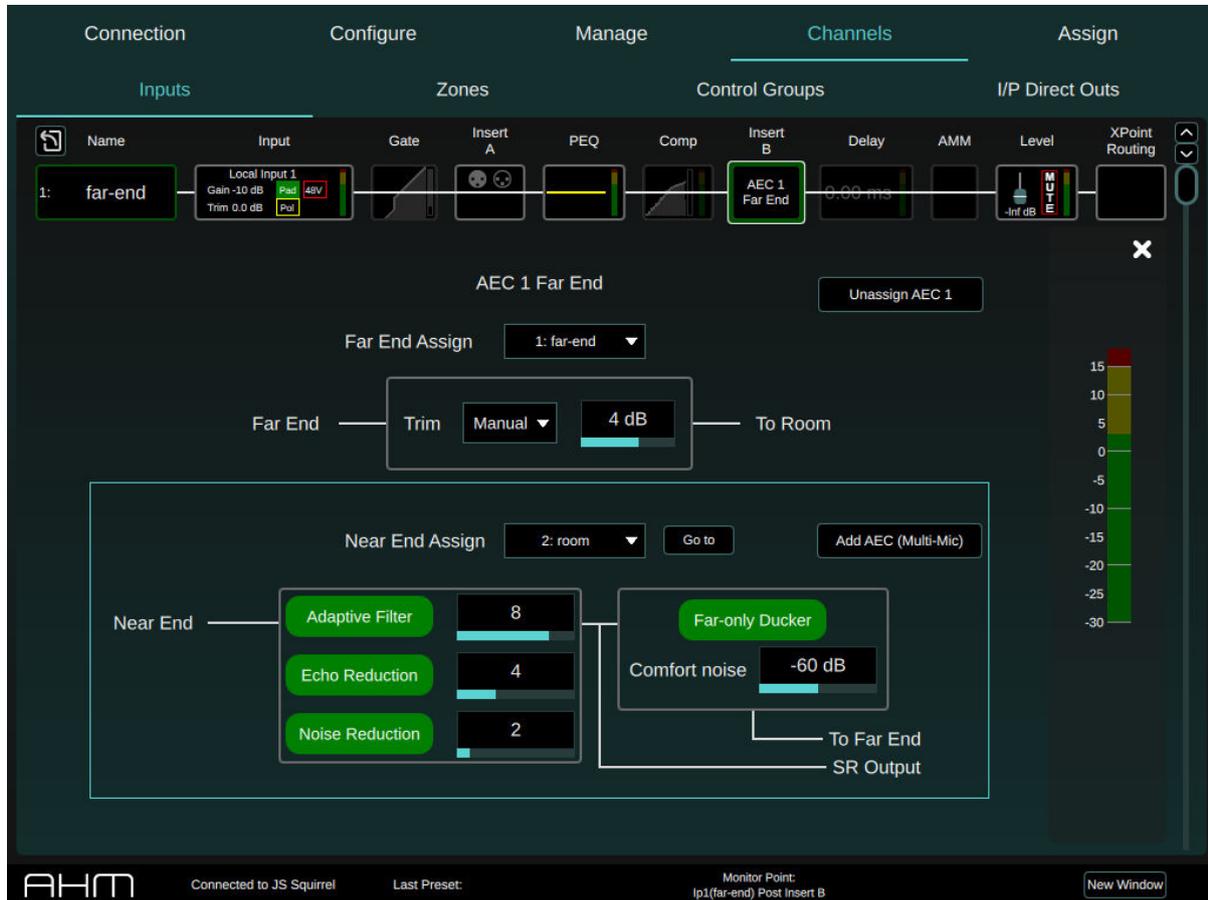
ALLEN & HEATH®

User Interface

Start with the far-end insert. Go to Insert B for the far-end Input channel, click on **Assign AEC (far-end)**, and choose one of 12 AEC processors to insert. AECs that are already in use will

be greyed out. The AEC settings and parameters are displayed.

ⓘ Requires AHM System Manager and firmware V1.10 or higher.



To assign the near-end, choose the source in the **Near End Assign** drop down in the screen above. Alternatively, go to Insert A for the near-end Input channel, and choose **Assign AEC (near-end)**.

ⓘ Note that the route from the far-end insert to the near-end speaker must not include any dynamic processing (compression or gating).

Click the **Go to** button in the lower panel to switch between the near-end and far-end insert screens. The far-end screen shows the AEC reference meter, and the near-end screen shows the AEC output meter (with echo cancellation applied).

ⓘ The AEC will not work if the echo seen by the near-end insert is higher in level than the far-end insert return. With good gain structure and appropriate local mic gain, this should not be a problem.

To set up a multi-mic system, in the near-end insert screen, click **Add AEC (Multi-Mic)**. You will then be able to select all linked AEC processors and assign their near-end Input channels as appropriate. Using an AEC processor for each microphone source, as opposed to a single AEC on a microphone mix (or conference system output) will result in the best overall performance.

ALLEN&HEATH®

For voice reinforcement purposes in the local room, it is recommended to use the SR output instead of the processed near-end Input channel. Go to the **XPoint routing** screen in the Zone, and adjust the level of the appropriate **AEC SR Return** channel. This method should offer the best trade-off between echo cancelling performance and voice reinforcement. Latency is introduced by the Adaptive Filter, comfort noise and gating are not present in the SR output.

Another method is to split the microphone source into a second Input channel for voice reinforcement use. This will give the best quality of voice reproduction, without the latency and artefacts of the AEC processor, but can be detrimental to echo cancelling performance.

Controls and parameters

Adjust the **Trim** so that the meter shows around 0dB with a typical far-end source. A low-level far-end signal will give poor AEC performance.

Choose **Auto** to enable the AGC and disable the manual trim.

Adaptive Filter enables and controls the speed of the Adaptive Filter. Set to the lowest value that gives acceptable performance.

Echo Reduction enables and controls a part of the NLP block that further reduces echo. Set to the lowest value that gives acceptable performance. A high value can cause distortion to the far-end receiver.

Noise Reduction enables and controls part of the NLP block that reduces constant background noise from the local room. Once again, set to the lowest value that gives a useful background noise reduction.

For better rejection of low frequency noises, particularly those sporadic in nature, use the near-end Input channel HPF.

Far-only Ducker enables an attenuation stage driven by the talk state detector. The near-end signal will be reduced when the the far-end only is talking.

Comfort Noise controls the level of the speech-shaped noise sent to the far-end caller. The noise cannot be turned completely off, because some noise improves the effectiveness of the NLP block.

Tech Insight – Talk-state Detector

Most AEC system block diagrams show at least a Double Talk Detector (DTD), and often one or more voice activity detectors (VAD). These have been combined into one block called the Talk-state Detector (TSD). The primary function of a DTD and hence this TSD is to determine whether signals picked up by the near-end microphone contain both near-end speech and echoes of the far-end signal, i.e. double-talk. The adaptive filter will diverge if an adaptation pass occurs when there is speech in the near-end room, so this must be avoided as far as possible.

The AHM implementation uses estimates for the statistical properties of the band-pass filtered near- and far-end signals to drive the TSD decisions.

Tech Insight – NLP

Although the adaptive filter is effective at identifying the room impulse response, some residual echo will remain. This is especially true when the room response changes, or when the system has only just been activated. For this reason, a residual echo suppression system is used to try to compensate for these scenarios, providing better overall performance.

The system uses a spectral subtraction technique where both the far-end and adaptive filter error signals are windowed and transformed into the frequency domain. This is an existing technique, with the main challenges being the minimisation of artefacts and distortion that can often result from the process, and the obtaining of a good estimate of the values to be subtracted.

Background noise reduction

Background noise can harm intelligibility and impede the residual prediction system. The background noise spectrum can be estimated by tracking the power spectrum minima. To prevent near-end speech from affecting the background noise estimate, the estimate is only updated when the near-end microphone is deemed to not be active, i.e. the short-term SNR estimate from the Talk-state Detector is low.

ALLEN & HEATH®

Tech Insight – Adaptive Filter

Frequency-domain implementation

The adaptive filter is a Finite Impulse Response (FIR) filter, and as the system impulse response increases in length, the filter length must also increase to match. For a typical AEC adaptive filter length of over 100ms, the computing power required to implement such a convolutional filter is extremely large. A technique for reducing the computational requirement is to apply the filter in the frequency domain. This is beneficial because convolution in the time domain is equivalent to multiplication in the frequency domain. This being the case, the cost of forward and inverse Fourier transforms is quickly overcome by the reduction in cost of applying the filter. The trade-off with this technique is that a transform of useful length automatically incurs a delay, which would not be present in a purely time-domain solution. Filter partitioning techniques allow this latency to be brought down to an acceptable level, while still giving most of the computational benefit.

NLMS adaptation

Normalised adaptation dynamically adjusts the filter update rate based on the amount of energy in the signal, on a per-frequency basis. Without normalisation the system is very sensitive to operating level; small signals are not

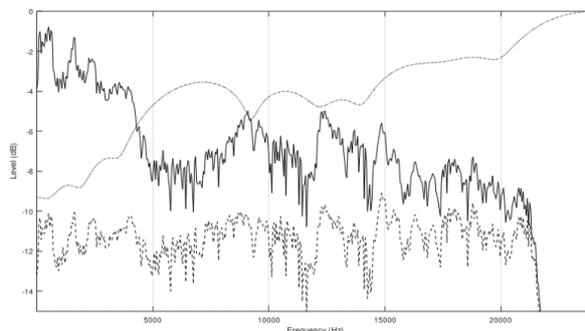
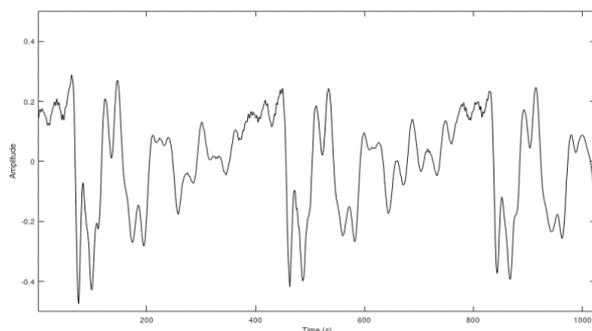
able to produce significant changes to the filter, and large signals can cause the filter to become unstable.

Signal whitening

Whitening is an established technique in adaptive filtering, where some prefiltering is applied to the input signals in order to give them a flatter spectrum. This is advantageous for the same reason that an adaptive filter will converge most quickly when the signal is white noise; the whiter the signal, the less likely it is to erroneously correlate with anything other than the true echo signal, thus the correlation in the filter update process will be far more likely to cause filter coefficients to move towards convergence, rather than away.

Since we need to adapt based on whitened signals, the output, or 'error' signal from this adaptive filter is not the echo-cancelled speech we need. As such, the coefficients generated from the whitened path need to be copied into a second filter, which is used to generate an echo estimate and produce the real AEC output.

The plots below show an example speech sample with corresponding frequency spectrum, and the frequency response of the resulting AR modelled whitening filter. It is clear from the whitened result that this process is effective in flattening the spectrum, providing a more ideal input for the adaptive filter. Also clear is that the whitened spectrum has less power; this has no impact on the system, due to the NLMS normalisation.



Left: Speech sample. Right: Speech sample spectrum (solid), whitening filter response (dotted), whitened spectrum (dashed).

For further information, application guides, and recommended products please visit

<https://www.allen-heath.com/installation/>

Don't hesitate to contact our Install team if you need assistance on which products to specify or if you have questions about an application.

ALLEN & HEATH®