XILICA APPLICATION GUIDE

Acoustic Echo Cancellation (AEC) →

This application guide is intended to serve as a reference to assist AV system programmers and field engineers in understanding the basic principles behind AEC and successfully deploying conferencing systems using AEC technology.

What is AEC?

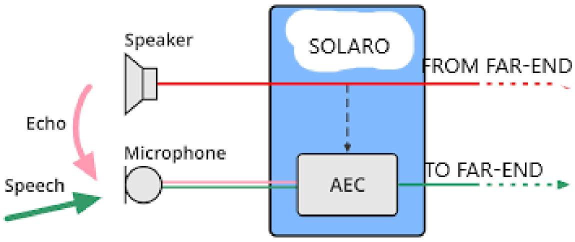
AEC is an algorithm in a digital signal processor (DSP), computer, or stand-alone hardware or software conferencing solution that removes unwanted echoes that can occur during a teleconference.

Why is AEC necessary?

In a conferencing environment, when the far-end caller is reproduced on near-end loudspeakers, that far-end audio will be picked up by the near-end microphone and transmitted back to the far-end. Because of the delay in transmission (whether over the phone or the Web) the audio returned to the far-end with arrive noticeably later, resulting in what are perceived as echoes.

The simplest real-world example might be a Teams, Zoom or Skype call on laptop using the computer's built-in speakers and microphone, as opposed to using earbuds or a headset. As you're hearing the remote party's voice on your computer speakers, that audio is also being picked up by the computer's microphone and sent back to the remote party. Without AEC at the near-end, that audio will be heard by the remote party as a very distracting echo.

How does AEC work?



The AEC at the near-end is always for the benefit of the far-end caller. As the far-end party's audio is received at the near-end conferencing system, the AEC algorithm takes a direct feed ("reference") of that far-end audio before it's locally reinforced on loudspeakers. It then compares that reference signal to what's being heard on the open near-end microphone. Any audio that's in the reference that's also heard on the near-end microphone gets cancelled from the signal by the AEC so that the audio sent to the far-end contains only the near-end talkers.

That's the simple explanation. There are actually a whole host of processes that make up a complete AEC conferencing solution. If you want to go into detail about these components, keep reading. If you just need to get a system up and running feel free to skip to the System Design & Signal-flow Best Practices section.

AEC PROCESSES IN DEPTH

Adaptative Filtering

As the far-end audio is being reproduced on loudspeakers in the near-end room, the near-end microphone is "listening" to the room and inputting that signal into the AEC processor. Using that signal, the AEC processor captures an image of the room in what's called a *finite impulse response* (FIR) filter. This FIR filter represents the acoustical properties of the room based how the room reflects and attenuates a given signal in that environment. This FIR filter can be applied to an audio signal to mimic what that signal would sound like if it were reproduced in that exact space.

So this FIR filter is applied to the incoming far-end's reference signal. Providing the near-end microphone was able to capture a pure image of the room (while it was quite and with little or no distortion of the signal paths), it creates a FIR filter that when applied to the far-end reference will result in a match between how the near-end microphone is hearing the far-end as it's reproduced in the room, and the far-end reference with the FIR filter applied.

With these two images captured, the AEC can then distinguish far-end audio as heard in the near-end microphone from other sounds (i.e. speech of the local meeting participants). Now we get to the actual cancellation in *acoustic echo cancellation*. Remember, we ultimately want to eliminate the far-end, as heard by the near-end microphone, from being transmitted back to the far-end.

With this reference information mimicking the actual near-end space, the AEC creates a replica of the echo and then filters the reference through the room impulse response estimate. This subtracts the replica echo from the microphone resulting in a reduction of the echo. The closer the echo replica to the real echo, the better the echo cancellation.

in effect applies an inverse of that audio signal to the signal being transmitted back to the far-end. The result is a subtraction of the far-end from that signal and, in theory, either silence or only the speech of the near-end party. This subtracting of a signal is called destructive interference and is the result of summing two waveforms that are an inverse or mirror image of each other. This is the same concept and process behind noise-cancelling headphones.

At this point you might think that we're done, and we may well be if the near-end room environment were perfectly static. But in the real-world, that space is actually constantly changing. As the near-end microphone is listening to the room so that the AEC can create the representative FIR filter, any change in microphone position, the number of people in the room, or rearrangement of any objects in the room effectively changes its acoustical properties and thus the work the AEC needs to do to successfully cancel the far-end from the microphone.

This is why there is a secondary process within the AEC called the *Adaptive Algorithm*.

Adaptive Algorithm

As the subtractive/cancellation process occurs within the AEC to cancel out the far-end, there is another downstream process occurring called the Adaptive Algorithm. The Adaptive Algorithm's job is to simply evaluate how successful the cancellation is of the far-end audio as heard in the near-end microphone.

If something changes in the environment and the result of the cancellation process contains left over artifacts, then the FIR filter needs to adapt to match the environmental changes. Based on what's left over from cancellation iterations, the Adaptive Algorithm uses that information to inform a commensurate update, or adaptation, to the FIR filter. That's why it's called an *adaptive* filter.

So, these two processes, the Adaptive Filter and the Adaptive Algorithm are constantly working together to first apply a filter to subtract far-end audio, then evaluate whether any signal remains, and then if necessary, roll changes based on the leftover artifacts back to the FIR filter. The ultimate goal is to get to what's called "convergence," which is basically the state where the AEC is successfully removing the reference material from the near-end microphone.

In a perfect scenario, all echo will be completely removed. In the real world though, imperfections and variables are introduced at various points along these signal paths (e.g., changes in the near-end room environment, end-user volume adjustments, amplifier and speaker distortion, etc.) and so another downstream process called Echo Suppression is introduced to catch residual echoes after the primary AEC adaptive filtering has occurred.

Echo Suppression

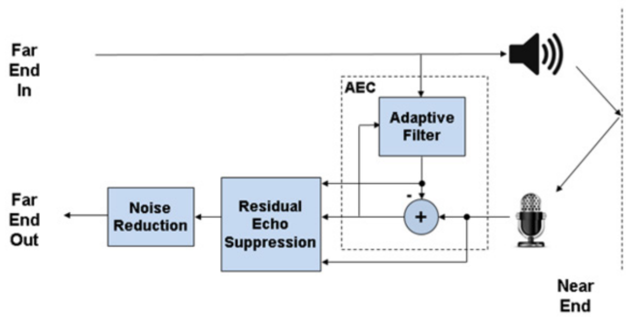
Echo Suppression is used as a secondary filter to catch residual echoes that slip by the primary Adaptive Filter. The NLP listens to the output of the Adaptive Filter and makes a determination based on probability as to whether the signal contains desired near-end audio, or the undesired far-end audio. Because a signal that contains both near-end and far-end content is all mixed together, the NLP operates like an intelligent multi-band noise gate. When it determines that undesirable far-end audio exists on a given band, the NLP attenuates that band to the degree of probability that the band likely contains far-end content.

Noise Reduction

Whether we're aware of it consciously or not, every space (aside from an anechoic chamber) has noise. This could be from HVAC, fans on AV or computer equipment, environmental noise like a road or highway outside a building, or any number of other sources that contribute to an environment's "noise floor." Thus, another process within a complete AEC solutions is Noise Reduction. The Noise Reduction process is downstream from NLP and simply seeks to cancel out mostly steady-state noise detected in the near-end microphone. The opposite of steady-state would, for example, be speech, where the signal is constantly varied (modulating in frequency and amplitude). The Noise Reduction algorithm captures an image of this noise and then applies subtractive filtering. Again, this is the same process behind noise cancelling headphones. Because noise reduction can have artifacts it should be dialed to taste by the AV system engineer.

Comfort Noise

Finally, after our near-end signal has traveled through all these processes with the over-arching goal of cancelling the unwanted far-end and any other noise, it's a bit ironic that the final process is actually one where we have the option of intentionally adding noise back into the signal path. After the original Adaptive Filter, NLP, and then additional noise reduction, the result should be a crystal clear near-end signal without any echoes heard by the far-end. Additionally, when neither the near or far-end is talking and the line is silent, it may in fact be so quiet that it sounds like the line has gone dead and the call has been dropped. So, as a subtle indication to callers that the line is in fact still active, a measured amount of "comfort noise" can be injected into the signal path. This is another parameter that should be dialed to taste by the AV system engineer.

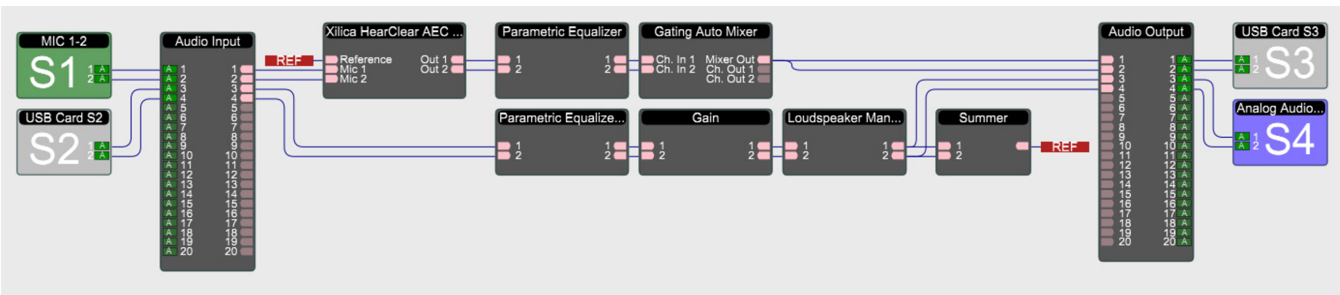


System Design & Signal-flow Best Practice

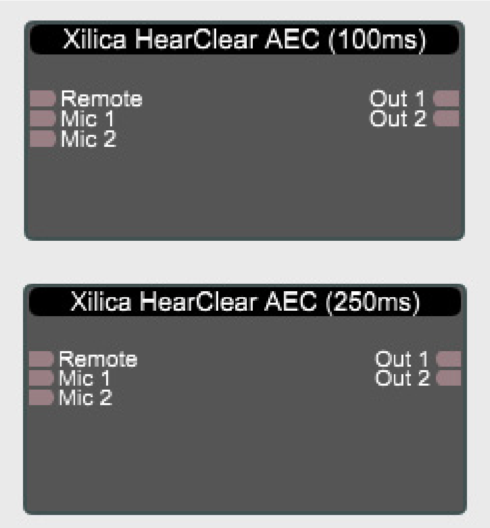
1. Place the AEC module at the frontend of the design. Any processing to the near-end microphone(s), including end-user volume and mute, should be downstream from the AEC module.
2. The AEC reference should be the same as the far-end signal that's sent to the near-end loudspeakers. Any processing that's done to the far-end, including end-user volume control and loudspeaker management, should be upstream from the reference signal.
3. If the near-end has program material that's locally reinforced on loudspeakers, this should be mixed with the far-end and included in the reference. *Simply, anything reinforced on the near-end loudspeakers that you don't wish to have transmitted to the far-end should be included in the AEC's reference.*

Commissioning an AEC System – Step-by-step

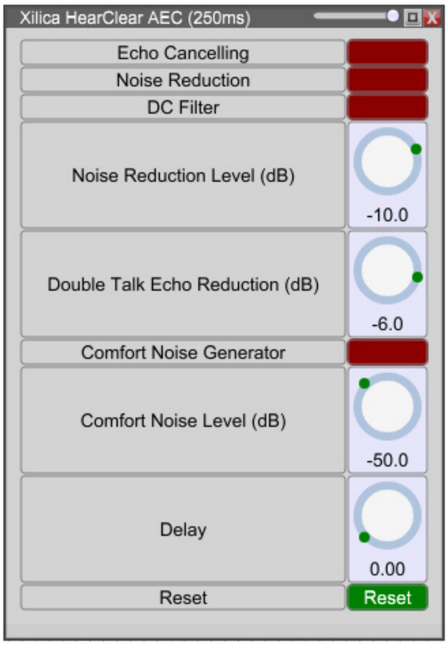
1. Turn off near-end power amplifiers.
2. Set the near-end microphone's input signal level for approximately -15 to -20 dB as shown on the meter inside the Audio Input block when talking at a normal level at a normal distance from the microphone.
3. Check the noise-floor of the room when no one is talking into the mic. It should be at most -40 dB or about 20 dB below normal active signal level. If there is a high level of steady-state noise in the room, consider using the Noise Reduction feature inside the AEC module.
4. Set any end-user volume controls at unity / 0 dB.
5. Have a far-end party call into the system and then set the gain level so that the signal meters at approximately -15 to -20 dB as shown on the meter inside the Audio Input block when talking at a normal level. It should be roughly the same as the near-end microphone level.
6. With the far-end party continuing to talk at a normal level, turn on the near-end power amplifiers and gradually raise the loudspeaker gain until the far-end talker is heard in the near-end room at a comfortable level, approximately 65 dB A-weighted.
7. Adjust the output gain of the near end microphone(s) inside the Audio Output block so that it's being heard at a normal and comfortable level for the far-end caller.



Xilica HearClear™ AEC Module



In the Component Library of the Design Screen in the Xilica Designer software there are two versions of the AEC module: 100ms and 250ms. These numbers represent "tail lengths" of a room's revertant qualities. For small rooms use the 100ms module and for larger, more reverberant, rooms use the 250ms module. The AEC algorithm can remove echoes from the near-end microphone that arrive up to 100ms or 250ms, depending on the module used. Note, the larger module will require double the DSP resources, and so half as many modules (4) can be utilized in one Xilica Solaro DSP.



Echo Cancelling

When enabled, the AEC will remove echoes for the far-end caller.

Noise Reduction

When enabled, it reduces the background noise in the near-end microphone signal by the amount selected on the corresponding dial.

DC Filter

When enabled, it will remove very low frequency content of the microphone signal and mitigate any asymmetry (direct current off-set) of the waveform that can interfere with the AEC process.

Double Talk Echo Reduction (dB):

A echo-suppression algorithm that reduces residual echo that exists downstream of the primary echo canceller. Note, a possible side-effect is reduced double-talk performance, so only use as much as is needed.

Comfort Noise Generator

When enabled and adjusted by the corresponding dial, it will insert noise into the feed returning to the far-end as a subtle indication that the connection is still live. Because the AEC and Noise Reduction algorithms can remove all noise from the signal, comfort noise is often added to simulate the natural background noise of a live call.

Delay

When enabled, it inserts a delay as specified by the corresponding dial to compensate for excessive proximity between the near-end loudspeakers and microphone.

Reset

Resets the AEC module. If AEC is deployed in a multi-purpose room with various conferencing configurations or locations of microphones, it can be useful to trigger an AEC reset when the room is physically reconfigured.

Other Tips & Best Practices

1. Optimize a conference room by installing appropriate acoustical treatment to dampen acoustical reflections that make AEC more difficult.
2. When designing the conferencing system try to create ample proximity between noise sources, including loudspeakers, and the near-end microphones.
3. Do not feed any near-end microphones to the AEC reference.
4. If using a conferencing codec, disable the AEC in the codes and rely solely on the AEC in the DSP.
5. Because program source levels can vary widely, consider using an automatic gain control (AGC) on these sources to help keep levels consistent.
6. If program source devices have output level control, set the DSP input for a loudest-case scenario to avoid possible clipping.
7. Keep near-end loudspeaker reinforcement levels conservative and close to a conversational level. Loud rooms make AEC challenging.

Troubleshooting

1. If the far-end hears audio distortion that sounds like an "underwater" effect, then it's likely due to a near-end microphone being fed to the reference.
2. If the far-end hears themselves or echoes:
   a. Check that the far-end isn't internally routed in the DSP and transmitted back to the far-end.
   b. Check that the input gain of the far-end / reference is approximately the same as the near-end microphones.
   c. Increase the level feeding the AEC's reference input. Put a gain block directly upstream from the AEC reference input and boost it in 3dB increments while the far-end caller listens for improvements.
   d. Check that the appropriate AEC module (100ms or 250ms) is being used per the size of the room. If echoes are still heard when using the 100ms module, increase to the 250ms module if system resources allow.
   e. Adjust the Double Talk Echo Reduction parameter in the AEC module to taste to attempt to cancel any residual echoes.
   f. If it's a particularly large space, adjust the Delay parameter in the AEC module to taste to compensate for a large proximity between loudspeakers and microphone.

Conclusion

Designing and deploying a successful audio conferencing system utilizing AEC can be challenging and confusing depending on the complexity of the system and the acoustical properties of the room. Follow the steps in this guide and always remember the most important concept in AEC: any audio heard by the near-end microphone(s) that you don't want transmitted to the far-end should be included in the AEC reference.