ASPI DIGITAL
THE SOUND OF DSP

How to Choose an Acoustic Echo Canceller

TABLE OF CONTENTS

INTRODUCTION

If you have ever heard clipped speech, acoustic echo, or screeching feedback during a conference call, you've experienced an audio system in need of an Acoustic Echo Canceller (AEC). An AEC is a component of a hands-free communication system that filters out echoes and minimizes distortion of the audio signal. An AEC greatly enhances audio quality, allows conferences to progress more smoothly and naturally, and prevents listener fatigue. An AEC solution that is poorly designed or inappropriate for the location will not provide these benefits and may even degrade audio quality. Many AEC solutions are available, ranging in price from dollars to thousands of dollars, and covering a broad range of quality and performance.

Today, hands-free communication systems are becoming more prevalent for teleconferencing, video conferencing, and distance learning. If you are designing, building, or purchasing a hands-free communication system, you'll want to consider the need for an AEC. This paper describes what an AEC does (in comparison to similar solutions) and then presents step-by-step guidelines for choosing the best AEC solution for your needs.

OVERVIEW OF AEC

Acoustic echo is most noticeable (and annoying) when delay is present in the transmission path. This happens primarily in long distance circuits or in systems using speech compression (such as video conferencing or digital cellular phones). Even though the echo might not be as annoying when there is no delay (as with short links between conference rooms in the same building or distance learning over fiber-optic cable), it is still intrusive and uncomfortable and can cause fatigue and "listener stress" for conference participants. Also, high-pitched squealing or "howling" can occur if the microphone is too close to the speaker or speaker volume is too high, whether or not there is transmission delay. The risk of howling may require reducing either the loudspeaker volume or the microphone sensitivity, which in turn may reduce the effectiveness of the conferencing system. All of these problems are corrected by an appropriate AEC solution.

Acoustic echo cancellers can be used in both narrow-band (3.5 kHz) and wide-band (7 kHz) conferencing systems. Narrow-band applications include teleconferencing and low bit-rate video conferencing. Wide-band applications include high quality teleconferencing and video conferencing, as well as distance learning. Users of wide-band conferencing systems should be particularly interested in using an AEC solution, as it will help them to reap the most benefit from the additional audio capabilities of their systems.

People at the far (or remote) end of the transmission path are the primary beneficiaries of an AEC. Installed at the near (or local) end, an AEC prevents the echo of the remote person's voice from being returned (echoed) to them through the audio system. People speaking on the same (local) end as the AEC should not notice the AEC if it is doing its job properly. Since the person on the far end hears better audio quality, the AEC enables the conversation to flow more smoothly, benefiting both parties.

In order for participants at both ends (near and far) to hold a full-duplex hands-free conversation, both ends must be equipped with an AEC.

Users of wide-band conferencing systems should be particularly interested in using an AEC solution, as it will help them to reap the most benefit from their system.

HOW AN AEC WORKS

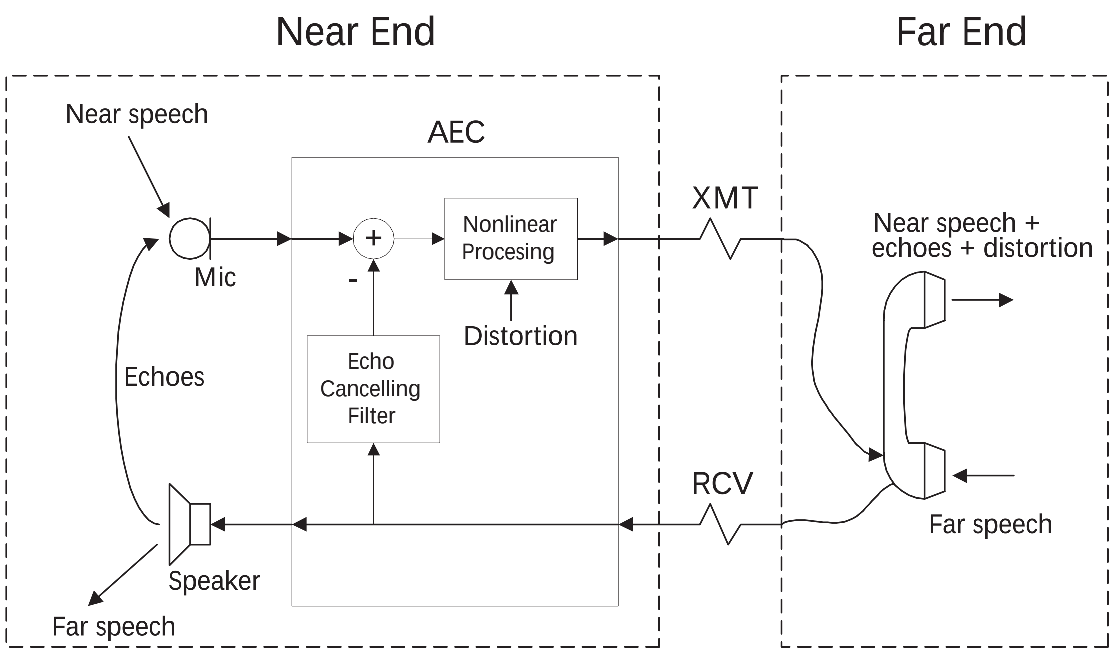

Figure 1. How an AEC Works. At the Near End, echoes of the far speech enter the microphone along with the Near Speech. The AEC filters out the echoes through an echo cancellation filter and nonlinear processing. In an AEC that is poorly designed, there may be residual echoes, as well as distortion added to the signal before transmission. (These effects are described later in the paper.) The Far Speech that travels through the receive path is not modified as it passes through the AEC.

WHY NOT USE A SPEAKERPHONE?

A speakerphone represents a simplistic solution to acoustic echo, because it allows only people at one end of the line to speak at any one time. In telecommunications terminology, a speakerphone is "half-duplex." It determines which side is active (or louder) by comparing the signal levels on both sides. It turns off the quieter side until the louder side is finished. Once one side has "captured" the circuit, most speakerphones do not permit any sort of interruption. This solution inhibits the natural flow of conversation and can hamper effective communication.

Speakerphones hamper completely natural communication flow.

ACOUSTIC ECHO VS. LINE ECHO

Line Echo Cancellers cancel echoes that occur in the phone line.

Line echo cancellation is a related but much simpler technology than AEC. Line echo cancellers are designed to eliminate reflections within a telephone line, rather than echoes from a room or auditorium. Acoustic echo cancellation is therefore a much more difficult problem. With line echo cancellation, there are generally only one or two noticeable reflections from telephone hybrids or impedance mismatches in the line. These echoes are usually delayed by less than 32 milliseconds, and do not change frequently, if at all. With acoustic echo cancellation, the echo path is very complex (dozens or hundreds of reflections), lasts 100-200 milliseconds, and can vary continuously during a conversation as people move around the room. While line echo cancellers may have smaller price tags, they don't perform under the conditions that acoustic echo cancellers must handle.

Steps to Choosing an AEC

Once you have determined that an AEC solution is needed for your application, you can use this four-step process to select the best solution for your needs.

1. Find AEC products with the form factor and features needed for your application.

2. Eliminate products that violate ITU G.167 convergence recommendations or the tail length requirements of the application.

3. Judge audio quality and state machine performance by comparative listening.

4. Weigh the performance, price, and convenience of each solution, and choose the one that will work best.

The rest of this paper discusses these steps in detail.

STEP 1:

FIND AEC SOLUTIONS WITH THE RIGHT FORM FACTOR AND FEATURES

There are several different AEC form factors to choose from. Be sure to choose one that best meets your needs.

By form factor we mean simply the form or packaging of the AEC solution. To a large extent, the products and systems that you build will determine which form factors are appropriate. As described below, original equipment manufacturers (OEMs), system integrators, and end users all have a range of choices available to them.

AEC FOR OEMS

Form factors for OEMs include pre-built modules, chipsets, and the licensing of algorithms. Which form factor is best depends in part on the volume of the product that is planned. While an off-the-shelf solution may be priced higher, it can decrease development costs and reduce time-to-market, which may be deciding factors for low-volume products.

- **Modules** are complete, off-the-shelf solutions. They are often suitable for moderate- to high-volume products and can speed the process of moving a product to market. They provide full functionality and quick integration into a design, as well as ensuring reliability and quality assurance. They can save a great deal of resources during the design process, and can provide a value-added feature to systems that may be used in a variety of applications. They may also offer such additional benefits as integrated A-D subsystems and other system support features.

- **Chipsets** are best for high to very high volume products. They allow tighter integration into a board, but require more effort and technical risk for the board design.

- **Algorithms** are best for very high volume products, especially embedded applications that are sensitive to size and power consumption. Algorithms provide the opportunity to use the processor for multiple tasks. They also can be ported to other platforms. While algorithms are the cheapest per unit at very high volumes, they require the most system integration work. This includes the supporting code, software interfacing, and integration with other resources.

AEC Solutions for Integrators and End Users

If you are a system integrator or end user of the hands-free communication system, you will normally choose a pre-built AEC solution. This again can take a number of forms. Which form factor is best will depend in part on the complexity of the system and the number of participants who will typically use the system.

- **AEC only (standalone AECs).** Standalone AECs are generally the least expensive, but require the integrator or customer to supply all external equipment, such as microphones, mixers and amplifiers.

- **Integrated AEC solutions.** This class of products may contain multiple microphone inputs, microphone mixers, sophisticated control structures to handle multiple connected "zones", and additional inputs / outputs for mixing or recording on tape.

- **AEC for videoconferencing.** Videoconferencing products may contain multiple inputs and outputs. They may also incorporate "phone add" modules to permit the addition of a two-wire conference (that is, allow callers from other locations to phone in to the video conference).

Features

Apart from the consideration of form factor, certain features may be desirable for your application. For example, wide bandwidth may be necessary for video or high quality audio conferencing. For integrated systems, the number and quality of microphones and speakers will be an issue. Automatic control of microphone and speaker levels may also be desirable. A graphical user interface (perhaps through a connection to a Windows computer) may be needed. These kinds of features are too varied to discuss in detail here. However, lack of such features may certainly be a reason to eliminate some AEC solutions from your selection process.

Product specifications should be viewed with common sense and a little caution. Different manufacturers assign different meanings to some classes of specifications. For example, a product that claims wide bandwidth may filter or pre-emphasize the signal to reduce low-frequency or high-frequency noise energy, effectively reducing the usable bandwidth. Specifications claiming unusually high cancellation figures or unusually short adaptation times should be examined carefully for consistency of interpretation.

Once you have found a number of AEC products that have the form factor and features for your application, you are ready to move on to Step 2.

STEP 2: FIND PRODUCTS THAT MEET G.167 AND TAIL LENGTH REQUIREMENTS

Meeting G.167 recommendations, is a good indication that the AEC filter has been implemented reasonably well. However, there are still other factors to consider.

The performance of an AEC product can initially be judged by two criteria:

- First, the product must be compatible with the ITU G.167 recommendation for AEC.
- Second, the AEC must have an adequate tail length for the environment where it will be used.

G.167 COMPLIANCE OR COMPATIBILITY

The ITU G.167 Recommendation for Acoustic Echo Controllers gives criteria for a number of performance characteristics typically listed on manufacturers' data sheets. These include such specifications as initial convergence time (or rate of convergence), and amount of cancellation. If an echo canceller has passed the G.167 AEC tests, this is a good indication that the filter algorithm (the actual echo cancelling filter) has been implemented reasonably well. It also means that the manufacturer has subjected the product to a series of standard tests, and that the specifications are likely based on valid experimental data. This makes the selection process easier, because it summarizes many different characteristics.

Compliance with recommendations for rate of convergence, or amount of cancellation does not guarantee that a given product will meet your needs; listening tests are necessary, as we will see in Step 3.

When an echo canceller has passed the G.167 tests (outlined in section 5 of the standard) the following specifications commonly found on data sheets have met the requirements of the standard in the room where the echo canceller was tested:

- Weighted Terminal Coupling Loss (or total cancellation)
- Initial Convergence Time (or convergence rate)
- Recovery Time After Echo Path Variation

Since most of the specifications found on data sheets are covered by G.167, it is not important to consider each of these specifications in detail. The manufacturer's equipment should have already been verified to meet the quality requirements of the standard. If the product exceeds any of the requirements, this may improve the audio quality to some degree. This improvement, however, will not be as significant as the effects of the tail length and state machine.

ROOM ACOUSTICS

G.167 testing is performed in real rooms. If the product meets the requirements in these rooms, it is compliant. A device that passed the G.167 tests in one room, however, might not be compliant in another. This is because the acoustics of all rooms are different. This flexibility allows manufacturers to design and test their products for certain types of rooms and claim compliance. But this also makes the customer responsible for determining whether the AEC will operate correctly in his or her particular environment. For example, an AEC solution that was designed to operate in an office may not work properly in a conference room.

If an AEC product is compliant in one room and not another, it is most likely due to a tail length that is too short for the second room or an inability to handle an increase in acoustic gain of the second room.

TAIL LENGTH

The tail length of an AEC defines the maximum echo delay that it can cancel. It is therefore critical to determine that the tail length will meet the requirements of the room where the AEC will be used. Since tail length is not specified by G.167, it must be evaluated separately.

Tail length is directly related to the reverberation time of the room. As the reverberation time increases, a longer tail length is needed. If the reverberation time is much longer than the tail length, a significant amount of the echo will remain audible. However, excess tail length will neither improve nor degrade the performance of the AEC. You should therefore determine the minimum tail length requirements for an application based on the typical acoustics of rooms where the product will be used. Any products that do not meet or exceed that tail length should be eliminated from consideration.

There are two main factors that affect the reverberation time of a room. They are room size, and the materials used to construct the walls and objects in the room. Most sound is absorbed when it strikes walls or other surfaces. If materials are used that absorb sound well, such as carpet, curtains, or acoustic tile, the reverberation will die out more quickly than if the room contains mostly reflective materials, such as hard wood, glass, or plaster. If a room is small, the sound waves will bounce off the walls more frequently and will be absorbed more quickly.

Tail length is important in making sure your AEC can handle any room size.

The following formula can be used to determine the necessary tail length for an environment. It relates the tail length to the room size and the number of cancelled reflections.

$$T = (N + 1) * d / c$$

- T is the tail length of the echo canceller (in seconds)
- N is the number of reflections cancelled
- d is the longest distance between walls (in units matching the speed of sound figure, c)
- c is the speed of sound (343 meters per second or 1125 feet per second at room temperature).

The equation assumes that both the microphone and the speaker are mounted on the same wall, which is the worst case in terms of the number of reflections that will be cancelled. In that case, N must be an odd integer because the even reflections travel away from the microphone.

For example, consider a 10x20x30 foot conference room with very reflective surfaces that requires five echoes to be cancelled. In such a room, a tail length of 6 * 30 / 1125 = 0.160 sec = 160 ms would be needed. Figure 2 shows how these reflections would travel back and forth across the room.

Testing and experience show that a tail length of approximately 200 ms is ideal for most situations. AECs with shorter tail lengths have difficulty operating in large conference rooms, classrooms, or acoustically live rooms. Longer tail lengths may be required for highly unusual circumstances, but may exhibit slightly slower convergence rates.

Howling Rejection

Howling rejection is important in cases where both parties are using hands-free communications systems. In these types of systems, it is very easy for the open microphones and loudspeakers to produce acoustic feedback, resulting in squealing tones, much like the feedback from a microphone in an auditorium. This obviously prevents any useful conversation. The most common way to avoid this problem is to implement howling rejection, typically done by shifting the frequency of the signal as it goes through the canceller. G.167 specifies a maximum frequency shift for howling rejection, but does not actually require that howling rejection be a part of an echo canceller. Generally, you should avoid any AEC solution that does not provide howling rejection.

After you have eliminated AEC solutions based on convergence rates, degree of cancellation, bandwidth, tail length and howling rejection, you will most likely still have a number of solutions under consideration. What remains are the characteristics that can only be evaluated by comparative listening. These factors make the most difference in how an AEC sounds.

After have completing steps 1 and 2, it is now time to perform comparative listening tests.

STEP 3:

PERFORM COMPARATIVE LISTENING TESTS

The best way to evaluate the audio quality of an AEC is by listening to it during a real conversation.

Ideally, an AEC lets speech signals pass through it without degrading them. This is actually the most difficult task in both designing and measuring the performance of an AEC. The main difficulty is in determining how the AEC sounds during double-talk (when people are talking at both ends of the line) and whether it harms sound quality by inaccurately determining whether it should be in the double-talk state.

The best way to evaluate the audio quality of an AEC is by listening to it during a real conversation.

The "state machine" is the logic within an AEC that decides if it should be in double-talk or another mode. The quality of the state machine drastically affects the audio quality of the system. Because the effects of the state machine are most noticeable with dynamic signals (such as those present during a real conversation), it is very difficult to quantitatively measure its performance. Consequently, the best way to evaluate the audio quality of an AEC is by listening to it during a real conversation.

HOW THE STATE MACHINE WORKS

State machines make the difference between a good echo canceller and a mediocre or bad echo canceller. Unfortunately, most tests on AECs are static: that is, the echo canceller remains in only one state during the test. For example, tests for initial convergence time (the time it takes the canceller to approximate the actual sound in the room) are done while the AEC is in receive mode. An excellent convergence rate does not guarantee that the system will be able to determine when to converge during a dynamic conversation. In other words, if the state machine is not robust, the other characteristics of the system will not compensate for this deficiency.

The state machine in an AEC chooses between one of the following four states:

- receive—only far speech is present
- transmit—only near speech is present
- double-talk—both far and near speech are present
- idle—no speech is present

The state machine must accurately choose between these modes for the AEC to operate properly. If it does not choose properly, speech may be distorted or the AEC may go out of convergence. Since the mode changes frequently during conversations (especially when more than two people are participating), state machine performance is extremely important.

The basic factors of the state machine's performance are:

1. The accuracy of determining the correct state
2. Impact on the signal if the wrong state is selected
3. How gracefully it switches between states

The two most critical states of the AEC are receive and double-talk. The receive state is the only opportunity for the echo canceller to converge correctly. It is also the time when the echoes are most noticeable (because they are not masked by speech from the near side). During the receive state, the echo canceller must converge rapidly and apply nonlinear processing to further reduce the echo. If the state machine does not detect a receive state correctly, it will not adapt to room acoustics and echoes will remain audible.

Double-talk is frequently mistaken for other states and has the most drastic effect on sound quality when it is incorrectly detected. If the state machine confuses double-talk for a receive state, it may decide to start converging. If it does, it will try to converge to the near talker's speech as well as the room response. This causes the canceller to go out of convergence. It may also apply nonlinear processing. This can result in excessive attenuation, noisy or scratchy speech, or half-duplex behavior (allowing only one end to transmit at a time).

When the state machine switches between states, there should be no audible transition. On a poorly implemented state machine, there could be noticeable changes in volume level, changes in background noise level, or even audible clicks as the state machine changes states. These would be especially noticeable during the beginning and end of pauses in conversation, or even between words. The state machine may even transition several times between modes, making an annoying series of clicks.

TRAINING NOISE

Poorly implemented state machines require training noise each time the AEC is powered up or is in a new environment.

Another sign of an inadequate state machine is the use of training noise at the initiation of a connection or on power-up. AEC adaptation can be roughly divided into two phases: large, rapid changes are required to adapt to major acoustical changes (such as moving to a new room); smaller changes are required to adapt to minor perturbations (such as people moving and doors opening). When an AEC is first powered up in a room or moved to a new location, it needs to adapt from its uninitialized state to learn the new acoustics of its surroundings. A good AEC will have a state machine that can adapt to this level of acoustical change with no difficulty, quickly and unobtrusively, by determining when it is in the receive state and adapting rapidly during that state. The state machine of a mediocre or poor AEC will not be as decisive or may even make incorrect decisions during this critical phase. As a result, the AEC will remain unconverged for an unacceptably long period of time or, in extreme cases, will never properly adapt. Some manufacturers of echo cancellers force the user to store the room characteristics after the initial convergence. This compensates for the fact that the echo canceller is not capable of converging quickly to major acoustical changes. You should avoid these echo cancellers!

Many mediocre AECs compensate for inadequacies in their state machine by restricting the rate of change and the amount of change that they allow in their adaptive filters. This prevents the AEC from going too far out of convergence by adapting too rapidly when it gets confused by a major perturbation, while allowing it to track relatively minor changes such as a door opening or slow movement of people in the room. In this scenario, the AEC must undergo a rapid training procedure to "learn the room" from its uninitialized state. Once trained, it can adapt to small acoustical changes, but major changes will require retraining. This training usually takes the form of a loud burst of noise or a sequence of tones, which the AEC uses to adapt to the gross acoustical characteristics of its environment.

Training noise is annoying and even unsettling and uncomfortable for users. It is also an indication that the AEC will have difficulty adapting to rapid changes in the acoustical environment.

The Listening Test

The listening test is the most important part of evaluating the AEC.

The listening test is the only way to evaluate the performance of the state machine, and is therefore the most important part of evaluating the AEC. A panel of several people should be chosen. If possible, the same people should evaluate all of the AEC solutions under consideration during a short period of time. These people should listen for the common problems listed below, as well as for the overall audio quality.

The most important part of the evaluation is on the opposite end from the echo canceller (the far or remote end). This is where the echo would be heard in the first place and most of the echo canceller's problems become evident. If the AEC is sold as part of a complete system (including microphones and speakers), some evaluation should also be done on the near end to ensure that all of the audio components are of good quality.

On the far end, either a handset or another AEC of the same type should be used. A listening test should not be performed with a half-duplex speakerphone or a different AEC on the other end, since it would not be clear which end had problems.

Ideally, the test environment should be the same (or at least similar) for all of the AECs, since room acoustics have such a large impact. If this cannot be arranged, at least consider the test environment differences in each case in the final decision. If possible, listen to the room acoustics with the echo canceller disabled so the effects of the different rooms can be compared.

Eight Things to Listen For

- **Residual Echo.** If there is excessive residual echo, the sound may have a hollow, distant quality or there may even be audible echoes. This is especially noticeable during the receive mode, when there is no near speech to mask the echo. If this is due to a short tail length, the residual echo may sound delayed.

- **Loss of Convergence.** When the AEC loses convergence, the result is an audible residual echo that could be louder than an echo heard with no echo canceller at all. This is generally caused when the state machine mistakes a double-talk situation for a receive state. If this happens, the AEC begins to adapt to the near talker's speech as well as the echo, and goes out of convergence.

The listening test is the most important part of evaluating the AEC.

- **Inability to Follow Changes in Acoustics.** Have the participant at the far (remote) end talk while a participant at the near (local) end is silent but walks back and forth in the pickup field of the microphone. The participant at the far end may notice a slight residual echo if there is excessive movement, but it should not be obtrusive. (Ideally there should be no noticeable echo during this condition). If the microphone is moved or the loudspeaker volume is changed, an echo may be heard briefly at the remote end, but the AEC should rapidly adapt to remove this.

- **Howling.** Pitched squealing noises may occur when both parties have hands-free systems with open speakers and microphones. This is caused by either a lack of howling rejection, or howling rejection that is not working properly.

- **Attenuated Speech During Double-talk.** To reduce the risk of howling or feedback during double-talk, the AEC may apply switch loss, which reduces the levels of the speech signals. Changes in volume levels during double-talk may also be caused if the state machine mistakes double-talk for a receive state and applies a different level of switch loss (attenuating the near signal to reduce the residual echo).

- **Half-duplex Behavior.** This is basically an extreme case of attenuated speech during double-talk. If one of the sides is attenuated so much as to become inaudible, then it would be impossible for speakers there to interrupt the other side.

- **Clipped or Noisy Speech During Double-talk.** Very harsh and annoying distortion can be added to the speech signal when nonlinear processing is applied during double-talk. The speech may be distorted beyond recognition. This occurs when the state machine recognizes a receive state during double-talk.

- **Audible State Transitions.** Audible changes in background noise level, clicks, or changes in overall volume levels may be noticeable during state transitions. This may even occur between words or short pauses in speech. This is caused by a state machine that switches between states too abruptly or too often.

Choose the best solution

By conducting the listening tests described in Step 3, you will have thoroughly evaluated the quality and performance of the AEC solutions. You are now in a position to measure these factors objectively, and to weigh them along with price, convenience, level of support, time-to-market, and whatever other factors may be important to you. By evaluating all of these factors, you can choose the best AEC solution for your needs.

About ASPI Digital

For over 17 years, ASPI Digital, a privately-held corporation headquartered in Atlanta, Georgia, has specialized in the development of new and innovative digital signal processing (DSP) products for real-time voice and audio systems. ASPI's echo cancellation and signal compression technologies are used by OEMs and other clients worldwide. ASPI is focusing its DSP and real-time system expertise on building solutions that improve the way people communicate remotely. ASPI also supplies its technology at the board and component level, enabling customers to incorporate these technologies into their product or system with a minimal amount of time and effort.

EchoFree™ EF400 Acoustic Echo Cancellation System

The EchoFree™ EF400 is a stand-alone high quality acoustic echo canceller designed for simple set-up and automatic operation. ASPI Digital's superior AEC algorithm permits its use in the most demanding environments. It is ideal for distance learning classrooms, conference rooms, courtrooms, and other audio or video conferencing applications where the acoustic conditions demand improved AEC performance.

GLOSSARY

Acoustic Gain A measure of the difference of the electrical level of the input signal from the microphone to the electrical level of the output signal sent to the speaker amplification system.

Bandwidth The frequency range the canceller passes without attenuation. For narrow-band (telephone) applications, this is defined as 300-3000 Hz. For wide-band applications (such as video conferencing) this is defined as 50-7000 Hz.

Convergence Rate The rate at which the echo canceller converges or matches the actual sound of the room. This is measured in dB per second. In order to meet the Initial Convergence Time of G.167, an echo canceller must have a convergence rate of at least 20 dB/sec.

Double-talk The state of the echo canceller when there is speech on both ends of the connection. This is the most difficult state to detect accurately, and most problems with audio quality occur during double-talk.

Far end The end of the connection opposite the echo canceller. The user on this end could be using a handset or another hands-free system. This is where the performance of the echo canceller is most noticeable. See also near end.

Half-duplex Behavior of most speakerphones, which prevents howling and acoustic echo by only allowing one party to talk at a time.

Howling Pitched, squealing tones that occur when hands-free systems at both ends of a connection have open speakers and microphones. This is similar to microphone feedback in an auditorium. Howling can cause damage to audio components if it is not attenuated.

Initial Convergence Time The time it takes for the echo canceller to converge to a specified amount of cancellation. In G.167, the canceller must have an Initial Convergence Time of 1 second, and must achieve 20 dB of cancellation within this time. Manufacturers may alternatively refer to the Convergence Rate of the echo canceller on their data sheet.

Near end The end of the connection where the echo canceller is located. This is where the echoes take place, although they are heard at the far end. See also far end.

Tail length The length of the filter which cancels echoes (measured in ms). The more reverberation a room has, the longer the tail length needs to be.