

All You Wanted to Know About Acoustic Echo Cancellation

Author: Puneet Gupta and Anil Kumar, Video Communications Team, Ittiam Systems

Abstract

While IP telephony is gaining popularity and is the direction towards the future, it comes with its own set of challenges. While there are multiple aspects like jitter and packet loss management that come into picture for combating the 'IP' (or network) part of the ecosystem, there is a whole cluster of pre/post processing challenges that one needs to face to provide an overall quality solution. One of the major challenges in IP telephony pre-processing lies in cancelling acoustic echo that marks the speaker phone operation. Although speaker to microphone coupling occurs in traditional PSTN telephony too, the echo is less annoying due to really low end to end latencies (<30 milliseconds round trip delay). In IP networks, however, the delay tends to be much higher (>100 milliseconds in a typical case), which makes acoustic echo much more noticeable and hence aggravates the problem due to echo.

Despite of the fact that the cancellation of acoustic echo forms a cornerstone of voice quality assessment by users, much is not understood about the dynamics involving an acoustic echo canceller (AEC). In this paper, we provide an introduction to various aspects of AEC including its generation and characteristics, primary quality metrics, challenges involved in implementation, the impact of hardware on AEC performance in a phone and some information on testing an AEC.

Introduction

Sources of Acoustic Echo

Acoustic echo is generated when the sound playing out of a speaker device is coupled back to the microphone via direct or indirect paths. Therefore, the talker at the remote end hears his / her own voice back after a tangible delay, and this is known as acoustic echo.

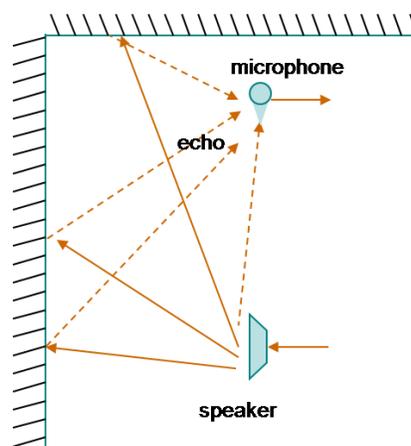


Figure 1: Illustration of Acoustic Echo Generation

The sources of coupling of speaker to microphone includes various paths:

1. Direct path between the speaker and microphone, if any
2. Reflections from the surface where the VoIP phone is kept
3. Reflections from the walls and other objects / people around the VoIP enabled phone
4. Coupling of sound via the physical enclosure of the phone, in form of vibrations from the chassis
5. Loopback modes in hardware audio codecs at the audio front end of the phone

Typically, echo characterizes speakerphone mode of operation, where the loudspeaker plays out the far end signal into the local ambience, and that gets coupled to the microphone. However, echo can also be heard in handset and headset conversations, since there are possible coupling paths even in those cases. The echo on the speakerphone however, is the most pervasive and difficult to cancel as compared to the other modes of operation.

Among the various interesting aspects of echo are the following:

1. Echo lingers for a finite and tangible duration after the remote end signal that generated the echo has been played out. Typically, one measures the duration which it takes for the echo to attenuate by around 60dB, and this is known as the tail length of the echo.
2. There are more than one echo paths at any given time
3. The echo paths can change dynamically with any change in ambience (including movement of people in the room), objects around the phone, weather and presence of moving objects in the room.
4. Different rooms or environments would have different echo characteristics. Typically, echo lingers longer in a larger hall or conference room (tail lengths of around 400 – 600 milliseconds) as compared to a smaller office (tail lengths of around 200 – 300 milliseconds).

Also refer to [ECDMYST] [ECNSCTL] for further information on sources of echo.

Primary Requirements for Acoustic Echo Removal

In a typical VoIP enabled phone, the typical requirements for an acoustic echo removal include (but not limited to) the following:

1. Adaptive filtering of echo
2. Ability to remove echo for variable tail lengths
3. Adaptability to be used in regular offices, small and large conference rooms as well as residential installations
4. Full duplex performance with ability for both ends to hear each other at all times during the call
5. Consistent performance for a wide range of speech signals and background noises
6. Ability to perform consistently for both narrow band (8kHz sampling rate) and wide band (16kHz sampling rate) signals

Besides the expectations on functionality, most realizations of VoIP enabled phones on an embedded system platform also require to achieve all this with minimal costs in terms of processor cycles and

memory requirements. This especially becomes a concern when echo removal is desired for long echo tail lengths or for wide band signals.

Challenges

Based on these requirements, the primary challenges in removing acoustic echo include:

1. Ensuring that there is no echo under any usage condition
2. Ensuring a full-duplex performance during double talk
3. Handling background noise
4. Handling unbalanced speech levels in transmit and receive paths
5. Ensuring consistent performance even during changes to echo path
6. Handling non linearity in echo path

Additionally, the echo removal techniques are also studied in relation with several standards for conformance. These include:

1. ITU-T G.167 – General Characteristics of International Telephone Connections and International Telephone Circuits – Acoustic Echo Controllers
2. ITU-T P.340 - Transmission characteristics and speech quality parameters of hands-free terminals
3. ITU-T P.831 - Subjective performance evaluation of network echo cancellers
4. ITU-T P.832 - Subjective performance evaluation of hands-free terminals
5. ITU-T O.42 - Equipment to measure nonlinear distortion using the 4-tone inter-modulation method

Although G.167 standard has been made obsolete by P.340, the former is still the *de facto* compliance that the VoIP industry adheres to.

Ingredients for an Echo Free Experience

In order to ensure that the users of VoIP enabled phones have an overall echo-free experience, there are three major aspects that need to be understood. These are:

1. **Echo Removal Algorithm:** This refers to the software method used for removing echo from the signal picked up by the microphone (near end signal). This algorithm forms the core of the echo removal technology, and the choice of a bad algorithm would result in poor user experience (echo leaks, half-duplex behavior, etc.)
2. **System Integration of the Echo Removal Module:** Implementing a good algorithm is necessary but not sufficient to ensure good perceptual quality from an echo perspective. Integration of this algorithm into the software system plays a critical role in the overall performance of the algorithm. This includes signal synchronization, gain calibration, and tuning of the algorithm for deployment.
3. **Hardware and Acoustics:** The whole intent of the echo remover is to counteract the acoustic coupling of the signal being played out with the one being captured by the microphone. A significant part of this coupling is contributed by the acoustic design of the enclosure, choice of microphone and

speakers, ADC/DAC components, etc. Inappropriate choice of parts or poor design of the hardware enclosure act as typical deterrents to quality even when the software system is first rate.

In the subsequent sections, we'll delve deeper into these three aspects of echo removal.

Echo Removal Algorithms

In order to get rid of acoustic echo, there are two primary categories of algorithms in use by most solutions. These are:

1. **Echo Suppression** – is based on the assumption that the communication is essentially half-duplex (only one end talking at a time). An echo suppressor simply detects whether the near end or the far end is talking, and closes the other path (like a switch). This technique, by design, is not conducive for ensuring full duplex performance in a double talk environment.
2. **Echo Cancellation** – is based on the principle of cancelling by using a filter to estimate the echo and removing it from the signal. Echo cancellation can be done in two modes:
 - a. Echo generated by reflections at Phone 2 and contained in the far end signal reaching Phone 1 can be cancelled by Phone 1
 - b. Echo generated by reflections at Phone 1 can be cancelled at Phone 1 and the far end signal reaching Phone 2 is echo-free (shown in Figure 3). This is a more standard mode of AEC deployment.

In this paper, we would consider only the acoustic echo cancellation (AEC) algorithms since they form a more generic case. An AEC algorithm can be envisioned as two primary parts: the adaptive filter algorithm and the AEC state machine.

The Adaptive Filter Algorithm

The adaptive filter forms the 'heart' of the AEC and determines how good the echo can be removed from the near end signal. There are several algorithms that have been tried in various realizations of an AEC, and have their own benefits and weaknesses. Some of these include:

1. LMS / NLMS (time or frequency domain)
2. De-correlation Filtering
3. Wavelet Based Cancellation
4. Sub-band filtering

A detailed literature survey of the existing echo cancellation algorithms, along with their mathematical formulation, is done by Rainer in [ECSURVY] and Frunze in [ECDMYST].

The AEC State Machine

All the adaptive algorithms assume that the near-end and far-end signals are mutually un-correlated. Without this assumption, the algorithm fails in the absence of double talk detection to inhibit convergence during periods of double talk. Hence, without a DTD, AEC works best for white noise signals only. However, since speech doesn't belong to this category of signals (mutually un-correlated), there is more that needs to be done. The most common approach is to ensure that the adaptation is halted when near end and far end signals are present at the same time. This is commonly known as **double-talk** condition, and lack of a robust double-talk detector can throw the adaptive algorithm off track. As a result, the stability and overall user experience of AEC depends on when the actual filter adaptation happens, and at what rate. This part of AEC constitutes the 'brain' of the algorithm. A good double-talk detection algorithm determines how full-duplex the AEC performance is, without compromising on how much of echo can be removed.

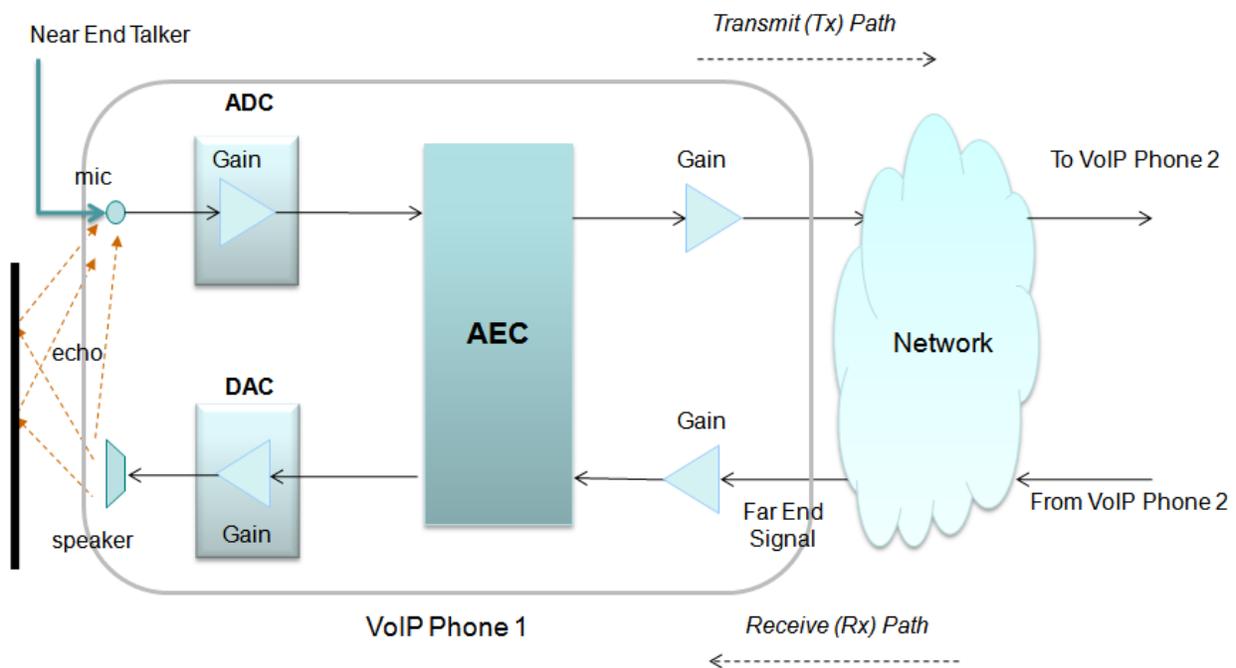


Figure 2: Acoustic Echo Cancellation in a typical VoIP Phone

System Integration

There are multiple aspects that have to be taken care of at the system level to ensure that the implemented AEC algorithm can well. In this section are listed some of the important system level considerations as far as the overall AEC performance is concerned.

1. **Gain Calibration:** In general, there are multiple analog and digital gains in the system. These gains are required to achieve required loudness for different signal paths. Distribution of gains in the path vis-à-vis AEC is very important as this will have significant effect on the performance of AEC. For example, analog gain is preferred to avoid quantization noise from digital domain, but may decrease the ERL. Similarly, some gain distributions can improve ERL, but may unbalance the levels of signal in

the transmit and receive path. Tuning of these gains (explained [later](#) in the paper) is therefore extremely important.

2. **Signal synchronization:** The fundamental premise on which most adaptive AEC algorithms work is the assumption of causality of echo with respect to the far end signal. What it means is that the echo is an effect of the far end signal, which is the cause. The AEC algorithm learns the echo path by observing the effect and the cause signals. Therefore, the reference and the response signals need to be synchronized in time before providing to the AEC algorithm. AEC filter may see non-causal signals if the echo signal is ahead of the reference signal. Similarly, if the echo signal is too much delayed with respect to the reference signal, then AEC would end up wasting some of the tail length in modeling the delay as a part of the echo path being modeled. Although the real acoustic response (from the speaker to the microphone) cannot be plagued by these problems, these effects might be encountered by AEC software due to capture and playback buffering done in a real-time VoIP enabled phone. Signal synchronization might also be compromised if scheduling is based on a soft RTOS operating system like Linux, which may result in an unpredictable scheduling of audio capture and playback paths. Recognizing the sources of and countering the loss of signal synchronization is a key aspect of AEC system design.

The Role of Hardware

Hardware design determines to a large extent the performance of a hands free phone (i.e., a phone with speaker -phone mode). Voice quality can suffer significantly due to poorly designed phone enclosure even when using the world's best AEC software. A few primary recommendations to ensure that hardware does not become the limiting factor in terms of AEC performance are listed in the following sections. For example: wideband speech codec support makes sense only when microphone and speaker are wideband. Similarly, if the microphone or speaker induces any non linearity in the echo path, then no linear model of the echo would be able to cancel the echo effectively.

Selecting the Right Microphone and Speaker

Speaker and microphone frequency responses and loudness ratings are subject to various standards across the globe. These standards define the recommendations for audio in digital telephony and telecommunication. The primary standards include:

1. TIA\EIA-810B - Transmission Requirements for Narrowband Voice over IP and Voice over PCM Digital Wireline Telephones
2. TIA\EIA 920 (North America) - Transmission Requirements for Wideband Digital Wireline Telephones
3. IEEE-1329.1999 (International)
4. ITU-T P.310-313, 340-342 (International)
5. IETS-300-245.2 (Europe) - Technical characteristics of telephony terminals; Part 2: PCM A-law handset telephony
6. IETS-300-245.3 (Europe) - Technical characteristics of telephony terminals Part 3: Pulse Code Modulation (PCM) A-law, loudspeaking and handsfree telephony

7. IETS-300-245.5 (Europe) - Technical characteristics for telephony terminals; Part 5: Wideband (7 kHz) handset telephony
8. IETS-300-245.6 (Europe) - Technical characteristics of telephony terminals; Part 6: Wideband (7 kHz), loudspeaking and handsfree telephony
9. TBR-8 (Europe) - Integrated Services Digital Network (ISDN); Telephony 3,1 kHz teleservice; Attachment requirements for handset terminals
10. TBR-10 (Europe) - Digital Enhanced Cordless Telecommunications (DECT); General terminal attachment requirements; Telephony applications
11. ITU-T O.42 (International) - equipment to measure nonlinear distortion using the 4-tone intermodulation method

Enclosure Design Recommendations for Speaker Phone Mode

Guidelines for the design of speaker phones more effective from an AEC perspective are listed below:

1. Make sure that the enclosure sits firmly on the base (for example: using rubber pads on the bottom of the enclosure) to help in reducing the speaker-to-microphone vibration coupling via the surface on which the phone sits.
2. Ensure that the microphone is housed in foam to increase ERL by reducing speaker-microphone coupling through direct path. Going one step ahead, one could also encase the microphone and surrounding foam in a separate housing within the enclosure.
3. Place the microphone sensor and speaker diaphragm at right angles if possible to reduce coupling
4. Build acoustical barriers between the speaker and microphone within the enclosure
5. Remove/re-design any circuitry that can cause EM pickup and remove all ground loops on the board to avoid electrical noise or hum.
6. Line up microphone with the hole in the phone enclosure for best pick up of the signal to help keep the gain in the microphone path minimal.
7. Operate the speaker, microphone, and amplifiers in their linear regime since AEC is a linear filtering algorithm and cannot accommodate any non linearity in the echo path.
8. Add sound absorbing material behind the speaker to damp out speaker resonances.

Some more useful guidelines are also outlined in [ECBSC] [LDSPKR] [HFDES]

Supplementary Features

Good enclosures are often quite expensive and challenging to design. Also, there is a possibility that external peripherals (microphones / speakers / amplifiers) are used in the end deployment scenario. Therefore, many of these guidelines do not get adhered in to in actual deployment. A typical way to address the issues that the deficiency in hardware designs is to include specialized signal processing components that can handle these deficiencies, in software. Some of these include:

1. Automatic Level Control
2. Noise Reduction / Cancellation
3. Speaker and Microphone Response Equalization
4. Comfort Noise Generation

5. Saturation Control

Of course, the trade off is in terms of increased costs on processing power and memory, along with additional engineering time. One should also keep in mind that the overall experience may become worse if any of these modules are not designed to take care of all use-cases.

Testing AEC

Testing of AECs is not a trivial task, primarily because the performance of an AEC in a phone is dependent on numerous environmental factors, which can change dynamically during the course of testing. The typical subjective tests that are conducted for an AEC, include:

1. Single Talk Tests
2. Double Talk Tests
3. Alternate Talk Tests
4. Tests with Background Noise
5. Tests with Music

Objective counterparts of some of these tests are also carried out by implementing simulation approximations to test recommendations like G.167 or P.340, either in software or hardware. Besides these, there are stress tests that system testers typically subject the AEC to. These include:

1. Extreme changes in the echo path, induced by
 - a. Moving an object / hand in from of the microphone / speaker
 - b. Covering the speaker with the palm for a few seconds and releasing it
 - c. Blocking the microphone input with a finger for a few seconds and releasing it
 - d. Changing the volume on the speaker path continuously up and/or down for multiple times
2. Sudden changes to the far end signal
 - a. Changing the test subject at far end from a male to female and vice versa
 - b. Changing the test subject at far end to a music player instead of a human talker
 - c. Extreme modulations in voice from far end (volume and / or pitch)

The objective test metrics that are typically noted or evaluated are:

1. Time for convergence in single talk
2. Depth of convergence during single talk
3. Worst case echo leakage during double talk
4. Worst case attenuation for near end signal during double talk
5. Speed of switching from far end only to near end only talk
6. Worst case level of residual echo after NLP action in single talk
7. Worst case divergence of adaptive filter in single talk during echo path changes

Calibration and Testing

As per various standards like TIA/EIA-810A, 920 (North America), IETS-300-245.2 and IETS-300-245.3, IETS-300-245.6, IETS-300-245.6 (Europe), there are specific requirements for transfer characteristics of telephony equipment. These standards specify the following:

1. Expected range of levels of loudness (loudness ratings for speaker and microphone i.e., for the receive and transmit paths respectively) for all usage modes (handset, loudspeaker, headset, etc.)
2. Expected nature of spectral response across the supported frequency band (a plot of amplitude with frequency), for both the speaker and microphone i.e., for the receive and transmit paths respectively.
3. Permissible levels of noise and distortion allowed relative to the signal level for both the speaker and microphone i.e., for the receive and transmit paths respectively.

These are based on numerous psychoacoustic measurements performed by the standards committees. By adhering to these recommendations, vendors can ensure that their transducers and enclosure of the VoIP enabled phone would have no audio quality concerns in absence of other impairments.

Calibration Steps

Gain Calibration

Once the AEC software is integrated into the software system, starts the task of ensuring whether it works as expected when put in the enclosure housing, interfaced with the selected transducers (microphone and speaker) and tested in a real world acoustic environment.

The first step in this direction is gain calibration. Tests are conducted to ensure that the transmit and receive path gains are in accordance with the Send Loudness Rating (SLR) and Receive Loudness Rating (RLR) expectations defined by the standards (listed above). In order to achieve the required loudness ratings, the system integrator needs to decide on how to distribute the gains among the different available gains (analog and / or digital) available in the system. A typical illustration of these gains is included in Figure 2. Besides these gains, AEC may have internal gain / attenuation logic of its own, and that needs to be taken into consideration as well.

To complicate matters further, it must be kept in mind that some of these gains are in the echo path, while others are not. The ones that are in the echo path have an impact on the ERL (Echo Return Loss) that the AEC needs to work under. A higher ERL means a friendlier environment for AEC to work in. Therefore, system designers should try to ensure two things while distributing gains amongst these gains:

1. No more than required gain should be added to the gains that lie in the echo path - this would help maintain a high ERL value
2. No more than required gain should be added digitally than required so that the SNR of the signal stays high (digital gains contribute to addition of digital noise)

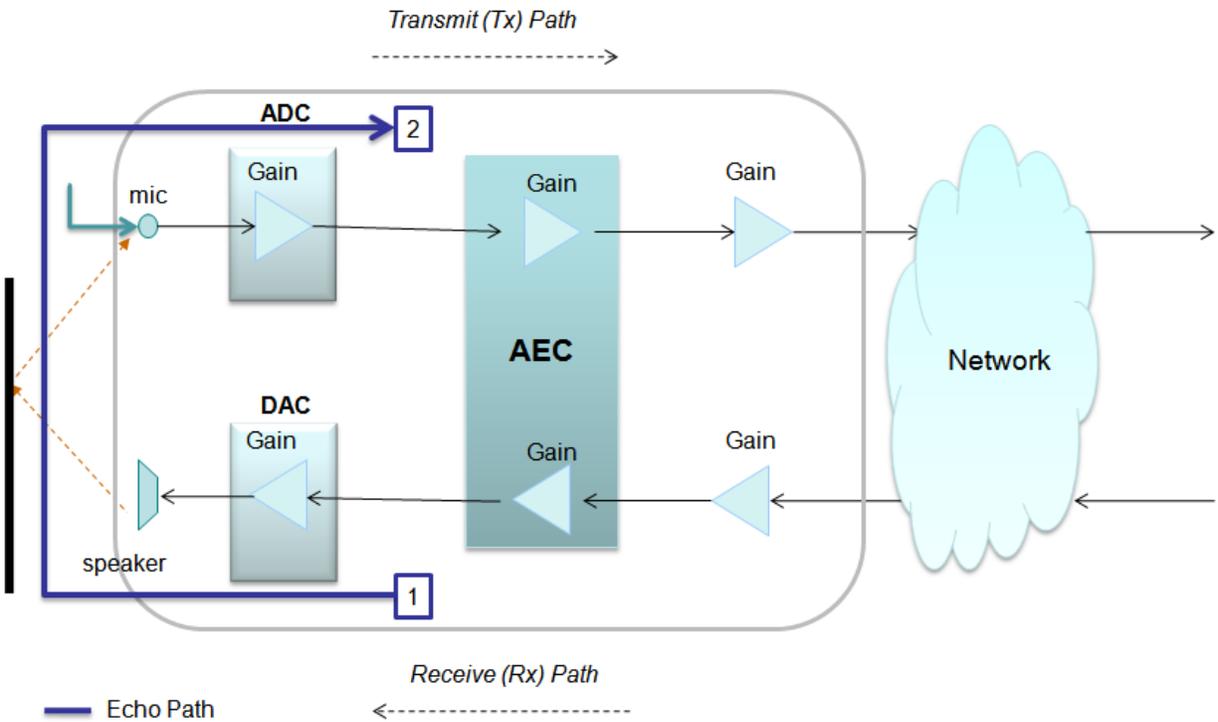


Figure 3: Illustration of echo path and system gains for a VoIP phone

ADC and DAC gains (Figure 3) are the most common forms of analog gains available to be added, but lie in the echo path. Therefore, the considerations listed above are mutually conflicting. It's the task of the system integrator to make sure that the trade offs are carefully balanced.

Further, VoIP phones need to operate at different loudness settings for the receive path – allowing the user the choice of volume for the incoming signal to be played out at. The TIA and ETSI recommendations listed above have guidelines on that the range of volumes too. The concerns that have been shared above in terms of gain calibration apply to all these volume settings, making the task of the integrator even more challenging. In general, louder the receive path signal, louder the echo and any gain applies to accomplish this loudness must take care of the following:

1. Far end signal should not get saturated in digital domain due to application of digital gain
2. If analog gain is applied, it must be ensured that the DAC is still operating in its linear regime
3. ERL should stay within acceptable limits i.e., gains applied in the echo path must be carefully controlled

Tuning of Spectral Response

Once the gains are balanced and tuned, the spectral characteristics of the microphone and speaker should be studied using standard signals like one sweeps, as specified by the TIA /ETSI standards. These tone sweeps help analyze the transfer characteristics of the transducers at different frequencies, and identify any issues with the same. In case transducers may have resonances at some frequencies, or

have a skewed frequency response (high peaks at some frequencies, and low troughs at others), then it does not reflect too well on the quality of sound they capture (microphone) or play out (speaker). In case the spectral response has any such issue, system integrators often employ the use of signal equalizers in hardware or software, which essentially perform FIR/IIR filtering operations on the signal just after the microphone captures it or just before the speaker plays it out. Once the response fits the recommended behavior, the phone can be considered fully calibrated.

Test Set Up Details

Testing of AEC enabled VoIP phones pose a challenge in terms of development as well as field testing. Some of the resources and facilities used typically in AEC test set ups are described here.

Silent Rooms

Quite often, the presence of noise in the ambient environment hinders the performance of the AEC. Note that even with the best cancellation technique, AEC might not be able to converge fully in presence of high background noise at near end, since distinguishing echo from near end noise becomes increasingly challenging with a degrading SNR (echo to noise). Therefore, many a times, the core AEC algorithm is first tested in an environment that is free of near end noise. Silent rooms provide such an environment. In absence of a near end talker, the level of sound picked up by a microphone placed in a silent room is under -60dBm0. Evaluation of echo path, offline training and tuning of AEC and designing algorithm updates to handle near end noise are typical activities that benefit from the presence of a silent room, not to mention approximate measurements for TIA810/920 and gain calibration.

Anechoic Chambers

Since reflections from walls and surfaces in the ambient environment is a primary source of echo, it is sometimes desirable to calibrate communication equipments in an environment that is free of echo. Anechoic chambers are shielded rooms that absorb acoustic echoes caused by internal reflections, besides providing extremely quiet conditions due to being isolated from external noise sources (just like a silent room). This helps the device under testing to not get influenced by external or internal acoustic interferences. As far as VoIP enabled phones are concerned, this means the ability to characterize the enclosure and transducers (microphones, speakers) in an ideal environment. Calibration exercises like TIA810A/TIA920 are best conducted in such an environment.

Facilities

There are various organizations that manufacture test equipment and have test facilities to carry out calibration and certification for standards like TIA810/920. Some of these include:

1. MWM Acoustics (www.mwmacoustics.com)
2. Microtronix (www.microtronix.ca)
3. Head Acoustics (www.headacoustics.com)
4. Bruel & Kjaer (www.bksv.com)
5. Hermaon Labs TI (www.hermonlabs.com)
6. GL Communications (www.gl.com)

Conclusion

This paper discusses the various challenges involved with designing a VoIP enabled phone as far as cancellation of acoustic echo is concerned. It is a problem that has many possible solutions and numerous aspects - from hardware to drivers to software - that impact the eventual performance of the phone. Perfecting the algorithm as well as the other aspects including calibration of system gains, synchronization of near and far end signals and enclosure design offer a lot of engineering challenges. It is extremely important that for any system that needs an AEC, a holistic design be done at the very beginning of the project, taking into account the various aspects mentioned above. Even if one of the pieces in this jigsaw puzzle does not fit well, it is quite likely that the whole system's performance would be jeopardized. The overall experience with AEC will be as good as the worst part of the puzzle. Engineering and re-engineering costs can also be avoided by taking a comprehensive early approach to AEC - often the most critical of the voice quality assessment aspects for all VoIP enabled phones.

References

- [ECDMYST] Echo Cancellation Demystified by Alexey Frunze, SPIRIT Corporation

- [ECBSC] [The Basics and Acoustic Echo Cancellation](#) by Bogdan Kosanovic, Texas Instruments

- [ECSURVY] [Echo Cancellation Techniques for Multimedia Applications - a Survey by Rainer Storn](#)

- [HFDES] [Solve distortion, echo return, and vibration in plastic hands-free designs](#) (by Dean Morgan, Zarlink Semiconductor)

- [LDSPKR] The Loudspeaker Design Cookbook, by Vance Dickason, Sixth Edition, 2000, ISBN-1-882580-33-8, Old Colony Sound Lab

- [ECNSCTL] Acoustic Echo and Noise Control: A Practical Approach by Eberhard Hänsler, Gerhard Schmidt, Published by Wiley-IEEE, 2004, ISBN 0471453463, 9780471453468