Voice Quality Enhancement
White Paper

Adaptive Digital Technologies, Inc.
June 30, 2009
Issue 1

Revision History

Issue	Description	Date	Engineer
1	Creation	6/30/2009	SDK

Table of Contents

1. Introduction 1
2. Voice Quality Degradations 1
   2.1 Echo 1
      2.1.1 Electrical Echo 2
      2.1.2 Acoustic Echo 2
   2.2 Reverberation 3
   2.3 Noise 3
   2.4 Variations in Signal Level 3
   2.5 Packet Loss 3
   2.6 Packet Delay and Jitter 3
   2.7 Frequency Response Variations 4
   2.8 Feedback/Howling 4
   2.9 Nonlinearity / Harmonic Distortion 4
   2.10 A Few Comments about Conferencing 4
3. Voice Quality Enhancement Techniques 5
   3.1 Echo Control 5
   3.2 Adaptive Beamforming 7
   3.3 Noise Reduction and Suppression 8
   3.4 Automatic Level Control 8
   3.5 Packet Loss Concealment 8
   3.6 Jitter Buffer 9
   3.7 Equalizer 9
   3.8 Adaptive Feedback Control 9
      3.8.1 Integrating Algorithms for Optimum Performance 9
4. Special Topics 10
   4.1 Cancelling Echo from the Network Side 10
   4.2 Beamforming / AEC Combination 12

1. Introduction

With so many different types of voice communication devices, equipment, and systems available today, it is increasingly important to employ Voice Quality Enhancement (VQE) techniques. At first that statement may seem odd. With all the technology that is built into modern communication equipment, one may ask “Why isn’t voice quality already good enough?” The answer to that question leads us to explore the types of voice quality degradations that exist in the various systems. And once we identify the voice quality degradations, we can discuss the voice quality enhancement methods.

The voice quality degradations (and associated enhancement techniques) that we will discuss are:

- Echo (Electrical and Acoustic)
- Reverberation
- Noise (electrical and background)
- Variations in Signal Level
- Packet Loss
- Packet Jitter
- Frequency Response Variations
- Feedback/Howling
- Nonlinearity / Harmonic Distortion

Our Voice Quality Enhancement software suite includes:

- Acoustic Echo Canceller
- Line Echo Canceller
- Network Echo Canceller
- Packet Echo Canceller
- Packet Echo Control
- Noise Reduction
- Noise Suppressor
- Automatic Level Control / Automatic Gain Control
- Packet Loss Concealment
- Jitter Buffer
- Equalizer
- Adaptive Feedback Control

2. Voice Quality Degradations

2.1 Echo

Echo is one of the most annoying degradation to voice quality in a telecommunication network. The reason is that when a person hears his or her own voice, it becomes very difficult to ignore the echo and continue speaking. The degree to which echo is a problem is a function of both the delay and the Talker Echo Loudness Rating (TELR). TELR is a measure of how loud the echo is compared with the loudness of the original speech. If a person’s echo is at a low level compared

to his or her speech, the echo is masked by the original speech. But if the echo is loud, it becomes a problem.

But that's not the whole story. If there is sufficient delay between the original speech and the echo, we become more sensitive to the echo even at lower signal levels or higher return losses. For example, if the echo is delayed by 5 milliseconds, a TELR of 20 dB may be tolerable, but if the delay is 100 milliseconds, the TELR needs to be greater than 40 dB in order for the echo to be tolerable.

2.1.1 Electrical Echo

The telephone network contains sources of electrical echo whenever a conversion is done between a 2-wire circuit and a 4-wire circuit. This is shown in figure 1. The most common device that connects to a 2-wire circuit is the standard analog telephone. The telephone is connected to the telephone central office by a pair of wires. This pair of wires carries speech signals from the telephone to the central office (receive) and from the central office to the telephone (transmit). At the central office, the 2-wire circuit is converted to a 4-wire circuit using a hybrid circuit. The resulting 4-wire circuit uses one pair of wires for each direction of transmission.

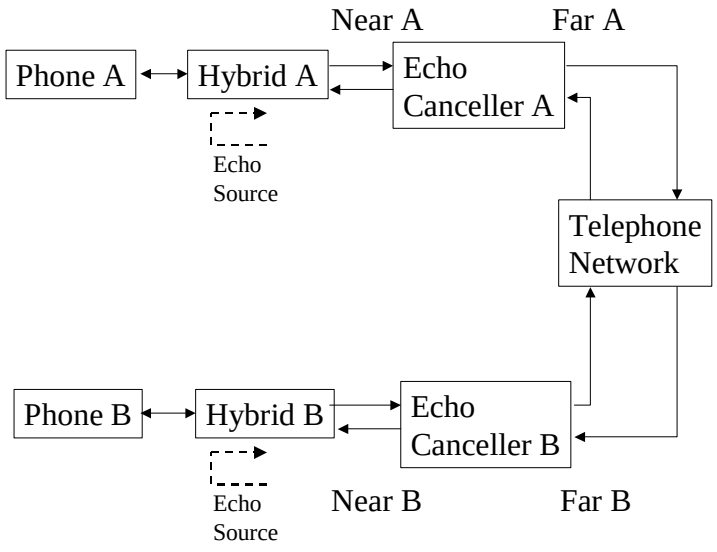

Figure 1 – Electrical Echo in the network

A hybrid circuit does its best to prevent from the signal being received via the 4-wire circuit from being reflected and echoed back. A typical hybrid circuit may provide 10-25 dB of isolation. But that is not sufficient for good voice quality unless the echo delay is very small.

2.1.2 Acoustic Echo

Acoustic echo is caused by feedback from a speaker to a microphone either directly or via reflection off of walls and objects. Although acoustic echo is more pronounced in hands-free phones due to speaker volume and microphone gain, it is also present in handsets and headsets to a lesser degree. Figure 2 depicts acoustic echo in a hands-free environment.

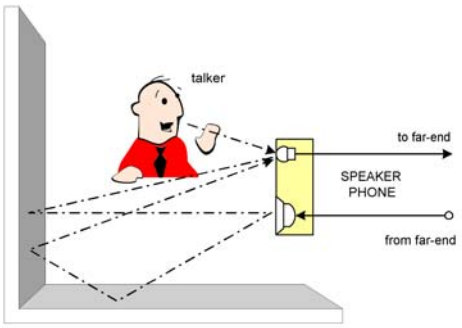

Figure 2 – Acoustic Echo in a Hands-Free Environment

2.2 Reverberation

Reverberation is similar to acoustic echo in that it is caused by reflections off of walls and objects. But in the case of reverberation, the person hearing the effect is not the one who is speaking. Reverberation becomes a problem when the delay of the reflective paths is long and the loudness of the reflections relative to the direct (mouth to microphone) path is high. Although this sounds similar to what we said about the circumstances under which echo becomes a problem, the acceptable delays are much shorter for reverberation.

Reverberation tends to become a problem when a person uses a hands-free phone and is not in close proximity to the microphone. This causes more loss in the direct path while the reflection path losses remain about the same.

2.3 Noise

Noise comes from two sources: analog circuitry and background (or environmental) noise. From an intelligibility standpoint, noise becomes a problem when there isn't sufficient signal-to-noise ratio (SNR). But even low level noise is a voice quality problem because it can be heard during quiet times when nobody is speaking.

2.4 Variations in Signal Level

Received signal level can vary due to gain and loss in analog circuitry. Much of the loss in an analog telephone circuit is due to the length of the wires that run between the telephone central office and the subscriber's premises. Loss occurs in hands-free situations is due to the distance between the person speaking and the microphone.

2.5 Packet Loss

Voice packets can be lost or arrive late in a VoIP network. This causes a brief but very noticible interrupt in speech.

2.6 Packet Delay and Jitter

In a packet network, the end-to-end transmission delay for any given packet can vary. It can vary so much in fact that packets arrive out of order. But for a voice communication system, the signal must be "played out" at a periodic rate and in the order in which they were transmitted – not necessarily in the order in which they were received.

2.7 Frequency Response Variations

Ideally we want the frequency response of a system to be flat over the entire desired frequency band. Sometimes this is not the case, and the voice quality suffers as a result.

2.8 Feedback/Howling

When both ends of a link are using a hands-free system, it is possible for an unstable feedback loop to occur due to the feedback between speaker and microphone at both ends. The problem is far worse of both ends are in proximity to each other and the microphones pick up the other end's speaker output. An echo canceller certainly helps with the first case if it engages before the howling condition begins. The second case is more difficult because of the crosstalk feedback.

Once howling begins, the cycle may not subside on its own. A user could mute the device or cover the microphone, but it is obviously better that he or she doesn't have to deal with this problem in the first place.

2.9 Nonlinearity / Harmonic Distortion

Nonlinearity and Harmonic Distortion can occur for many reasons. In the analog domain it is caused by signal compression and saturation, and analog-to-digital and digital-to-analog Analog converter inaccuracy. In the digital world it can be caused by overflow/saturation, and speech compression techniques, and "silence compression" techniques (also known as Discontinuous Transmission or DTX). People are particularly sensitive to this type of distortion.

2.10 A Few Comments about Conferencing

Distortion can be even more of an issue when dealing with conference calls. Let's take an example of a 20-person conference call in which each person dials into the conference on his or her own phone line. When a single person (person #1, for example) has uncancelled echo on his or her line, everybody else experiences echo whenever persons #2 through #20 speak. Not only is it annoying, but the process of determining the offending line is not intuitively obvious to the average person. It is actually the person who does NOT hear echo whose line is the problem. If two lines have uncancelled echo, it's virtually impossible to isolate the offending lines without having people successively hang up until the problem goes away.

High-density conferencing multiplies this problem further in that the probability that one party's line will have echo increases as the number of conference parties increases.

Noise can also be especially problematic in conferencing systems because the noise from each input is added together. When one or more parties are using a hands-free phone, especially a hands-free cell phone in a noisy environment like an automobile, a conference's voice quality can become much worse.

Variations in signal level are more noticible in a conference call for two reasons. First, the listeners must adapt back and forth between lower level signals and higher level signals. Second, some conferencing equipment identifies "dominant speakers" and only includes those speakers in the composite (summed) signal. Lower level signals will not have a fair chance at being identified as dominant compared with higher level speakers. They may even have a disadvantage when background noise becomes large.

3. Voice Quality Enhancement Techniques

Thanks to high-speed digital signal processors and some clever algorithms, we can counteract much of the distortion. We list again the types of degradation, but this time along side the voice quality enhancement algorithms that counteract the distortion

Degredation	**Enhancement Algorithm(s)**
Electrical Echo	Line Echo Canceller
Acoustic Echo	Acoustic Echo Canceller, Acoustic Echo Suppressor, Adaptive Beamforming
Reverberation	Adaptive Beamforming
Noise	Noise Reduction, Noise Suppression, Adaptive Beamforming
Variations in Signal Level	Automatic Level Control
Packet Loss	Packet Loss Concealment
Packet Delay and Jitter	Jitter Buffer
Frequency Response Variation	Equalizer
Feedback/Howling	Adaptive Feedback Control
Nonlinearity / Harmonic Distortion	That's a tough one. Suppression

3.1 Echo Control

Figure 3 is a generic block diagram of an echo control algorithm. It includes functionality that may be found in line and acoustic echo cancellers and suppressors.

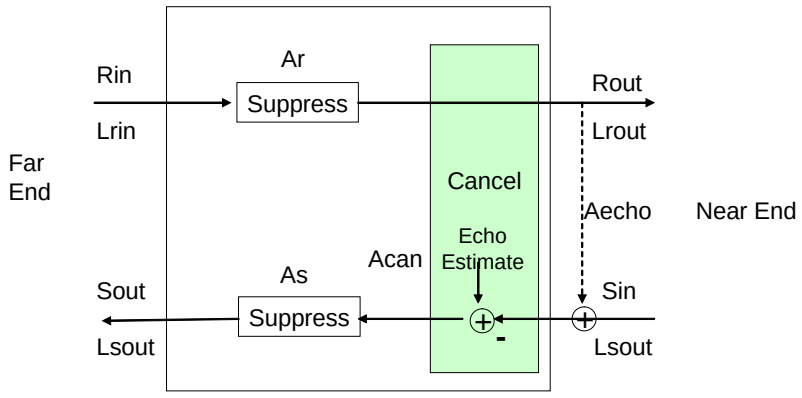

Figure 3 – Echo Control

The left side is designated the Far End. The right side, the side where the echo takes place, is designated the near end. The echo occurs at the near end, but it is the person at the far end that hears echo of his/her own voice. The far end to near end direction is designated the receive direction, and the near end to far end direction is designated the send path.

The levels at various points are designated Lxx. For example, the level at the receive input is designated Lrin. There are suppressors that operate in both the receive and send paths. The one and only canceller operates in the send path. The attenuation achieved by each of these blocks is designated Axx where xx is “t” for transmit, “r” for receive path, and “echo” for canceller. The amount of attenuation is not, in general, constant. One additional source of attenuation is shown – the echo source itself. The amount of attenuation of the echo source is designated Aecho.

The difference between a suppressor and a canceller is that the canceller removes echo without affecting the desired input signal. A suppressor applies loss to the entire signal.

The Send Input (Sin) of an echo canceller contains both echo and near end speech. The near end speech should travel through the send path with as little attenuation as possible. But the echo should be removed as best possible.

So, the canceller attempts to remove echo, but leaves the near end speech intact. The send side suppressor attenuates both the echo and the near end speech.

One might ask, why use supperssors in the first place? The answer is even the best canceller algorithms do not remove enough echo. There is always a small amount of “residual echo” that gets through. This is partly due to the fact that impairments like noise and nonlinearity interfere with the adaptive filter. Furthermore, near end speech interferes. So, the send side suppressor, sometimes known as a Nonlinear Processor (NLP), attenuates the residual echo even further.

One might also ask, why use a suppressor in the receive direction. There’s no echo at the Far End, and if there is, it is the responsibility of the Far End echo canceller to take care of it. The answer is that any attenuation in either path will improve the overall Talker Echo Loudness Rating (TELR), which is the figure of merit that affects voice quality.

TELR is the sum of all the attenuation sources in the path:

TELR = Ar + Aecho + Acan + As.

From this equation, we see that any source of attenuation improves voice quality from the point of view of the Far End speaker. But this only applies in the single-talk case, the situation where the Far End speaker is talking and the Near End speaker is not. When the near end speaker is talking, the send suppressor may attenuate his or her voice, which can degrade voice quality perceived by the Far End speaker. Similarly, when the Far End speaker is talking, the receive suppressor may attenuate his or her voice, which can degrade voice quality perceive by the Near End speaker.

The trick is to balance attenuation based upon the Aecho the canceller attenuation Acan, and the talk states (who is talking.) This is easier said than done.

As stated earlier, this is a generic description of echo control, where the term echo control covers line and acoustic cancellation and suppression. Differents of types echo control algorithms use the various algorithm components differently.

You can view echo control as a spectrum as seen in figure 4. At one end is a suppressor that does no cancellation. At the other end is a canceller that does no suppresson. The suppressor must rely completely upon attenuation. This results in a more half-duplex conversation. The suppressor must suppress one direction of speech nearly completely when both parties are

speaking. A canceller that uses no suppression achieves full-duplex operation because it does not rely at all on suppression. It is therefore subject to bleed-through of residual echo.

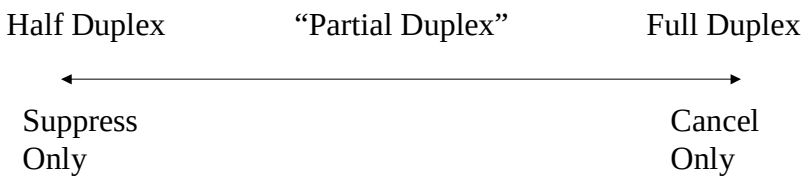

Figure 4 – The Spectrum of Echo Control Solutions

Lying between suppress and cancel are hybrids that employ both suppression and cancellation to varying degrees. More reliance on suppression results in a more half-duplex conversation. Less reliance on suppression results in a more full-duplex conversation. As stated earlier, the trick is a delicate balancing act.

3.2 Adaptive Beamforming

Beamforming is the process of combining the inputs from an array of microphones in such a way as to change the gain pattern of the array. The goal is to direct the maximum gain in the direction of the desired signal, as shown in figure 5. Since noise and reverberation signals tend to come from all directions, the desired signal is amplified while the noise and reverberation signals are not. This increases the overall signal to noise and signal to distortion ratio.

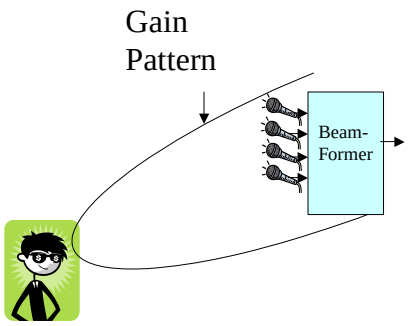

Figure 5 – Acoustic Beamforming

Adaptive Beamforming goes one step further by continuously identifying the location of the desired source and adapting the microphone array's direction to achieve maximum improvement dynamically.

Beamforming has an added benefit for hands-free scenarios in that the beam can be directed away from the echo source (the loudspeaker). It is even possible to place a null in the microphone array's receive pattern in the direction of the loudspeaker.

3.3 Noise Reduction and Suppression

One might ask, what is the difference between Noise Reduction and Noise Suppression? The difference is similar to the difference between echo cancellation and echo suppression. Noise Reduction is able to reduce the level of noise both during active speech and during quiet periods. Noise suppression is only able to reduce noise during quiet periods. Noise reduction therefore improves signal-to-noise ratio whereas noise suppression does not.

Noise Reduction is usually done by modeling the noise characteristics in the frequency domain, and performing spectral subtraction to remove the noise. Noise suppression is done by attenuating the signal during quiet periods only.

3.4 Automatic Level Control

Automatic Level Control attempts to maintain a constant output signal level regardless (within some limits) of the input signal level. In other words, it attempts to achieve a comfortable listening level. It may apply gain or loss to achieve this goal. It should be smart enough to amplify voice but not amplify noise when no voice is present.

3.5 Packet Loss Concealment

When one or more voice packets is not received, the information for a segment of speech is lost. But due to the short term redundancy in speech signals, it is possible to fill in the gap by looking

back at recent past speech and replicating it to cover the missing portion. This can be surprisingly effective for a single packet loss, but loses its effectiveness when too many subsequent packets are lost because the redundancy in speech is short term, and multiple replications sound artifical and. After a while, the replicated speech is no longer representative of the missing speech.

3.6 Jitter Buffer

A Jitter Buffer maintains a buffer of packets. The intent is to cover the timing jitter in packet reception. If we define the packet transmission delay to be the time between when a packet is transmitted and the time it is received, the packet jitter is the expected difference between the minimum packet delay and the maximum packet delay. In order to play out packets at the right time without missing any due to late packets, the jitter buffer size most be such that it contains enough packets to cover the expected jitter.

The Jitter buffer collects packets and starts playing them out when it has enough to cover the jitter. Although packets may arrive in an aperiodic fashion and in bursts, the jitter buffer outputs packets at a periodic rate as is needed for voice playout.

The Jitter Buffer is also able to resequence packets when they arrive out of order. It uses sequence numbers and time stamps that are embedded in the packets to accomplish this.

When a packet is delayed beyond the maximum jitter, packet loss concealment should be performed to minimize the voice quality effect of the lost information.

3.7 Equalizer

If a device has a frequency response that isn't flat, an equalizer can compensate by placing a filter in the system whose frequency response is such that when it is combined with the device's devices frequency response, the result is flat.

Of course, this requires knowledge of the device's frequency response. If the response is consistent from unit to unit, the same equalizer characteristics can be used for all units. If there is variation, the equalizer can be programmed at manufacturing time based upon measured frequency response.

3.8 Adaptive Feedback Control

The Adaptive Feedback Control quickly identifies and removes howling that is caused by an unstable feedback loop in a voice communication system. It is generally used when harsh acoustic conditions are expected. The Adaptive Feedback Control algorithm tends to be used as backup measure in case echo cancellation does not take effect in time to prevent the feedback from starting.

3.8.1 Integrating Algorithms for Optimum Performance

We have discussed many individual voice quality enhancement algorithms at a very high level. There is plenty of detail that goes into designing good algorithms. But the problem doesn't end there. These algorithms are dynamic algorithms. Putting them together properly into a system is crucial to achieving good voice quality.

In fact, for ideal performance, some algorithms should ideally operate "inside" other algorithms. For example, AGC and Noise Reduction work best when they operate on signals internal to echo cancellers. In order to facilate this, some of our algorithms are actually packages of algorithms

that are pre-integrated. Others have multiple APIs that allow the insertion of other algorithms into the data flow.

Figure 5 shows an example of how multiple algorithms can be integrated together with an echo canceller to achieve optimum performance.

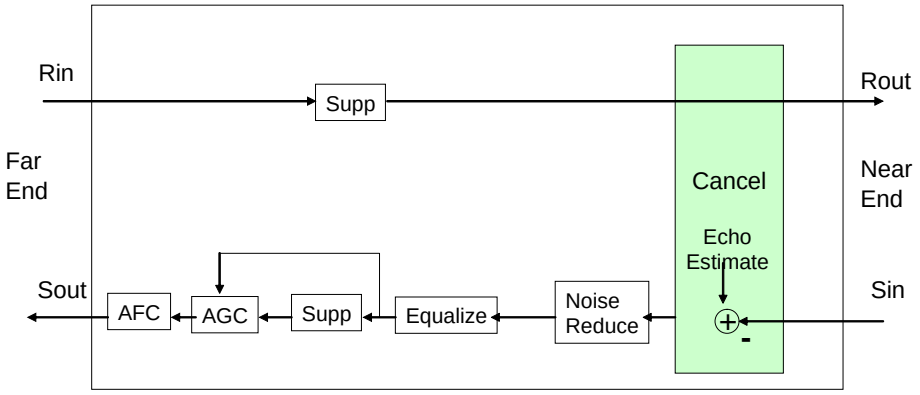

Figure 5 – Example - Integrated Algorithm

4. Special Topics

If we still have your attention this far into this paper, you are either a serious player in this space or you have way too much time on your hands. In either case, we'd like to present some voice quality issues and solutions that are not for the casual designer. But if you solve these issues, you will be one of the first to do so (at the time of this writing), and you will differentiate your product those of your competitors.

4.1 Cancelling Echo from the Network Side

According to well-established conventional wisdom, echo cancellation, whether electrical or acoustic, is best done as "close to" the source of the echo as possible. It is important that the echo canceller's reference signal (receive signal) reflect as closely as possible the signal that is being echoed and returned to the echo canceller's send input. Under these conditions, the echo canceller can best model the echo path.

But if the receive signal is modified after it leaves the echo canceller on the way to the speaker or during the return path from the microphone to the send input of the echo canceller, any such modification is done without the knowledge of echo canceller and can therefore be detrimental to the operation of the echo canceller. At best, the operation could be simple delay for which the

echo canceller is not properly compensated. Worse is a linear operation, which would spread the echo path impulse response, but still not so bad. Even worse is a time varying or nonlinear component in the echo path. A nonlinearity is not handled well by an echo canceller, and a time varying echo characteristic will cause the echo canceller to constantly try to adapt to changing conditions.

Having said all that, there are times when it is advantageous to perform echo cancellation at the network end of the link rather than the user end of the link. One reason for doing this is that there could be access devices (phones, ATAs, handsets) that are designed with sufficient echo control for one type of system delay, but when used in conjunction with a second system that incurs additional delay, the echo control is no longer sufficient. One example, shown in figure 4, is a DECT handset may have sufficient echo control when its base station is connected directly to the PSTN. But what happens if the DECT base station is instead connected to a VoIP gateway, causing tens of milliseconds of additional delay. If you recall from section 2.1 of this white paper, longer delay requires better echo control (better TELR). But the DECT handsets are already designed and deployed. Something has to handle the added requirement, and it's easier to put that responsibility into common equipment rather than replacing all the DECT handsets that are fielded.

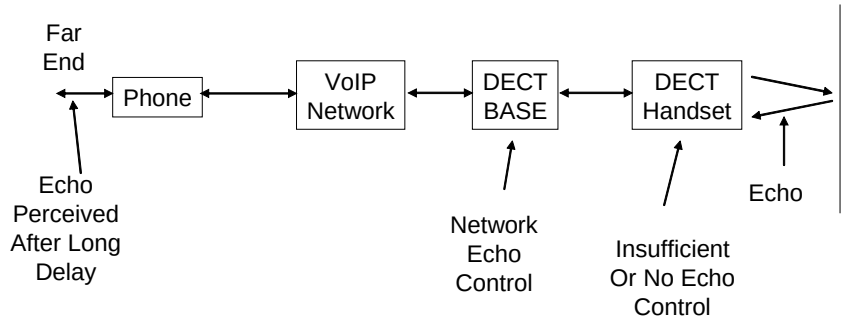

Figure 4 – Example – Echo Control done at network side in DECT Base Station

A second example is an IP conference server. A similar situation could arise in which there a device (phone) on a conference call that has insufficient echo control for the added delay incurred by IP conferencing. The problem is magnified for the reasons mentioned previously with respect to conferencing. In order to provide the best possible voice quality in light of this issue, it behooves the manufacturer of the IP conferencing equipment to put echo control into their equipment.

But when the echo control is placed in the network, the path between the echo canceller and speaker and back from the microphone to the echo canceller may indeed be modified by delay, bit or packet errors, and speech compression in both directions. To say that these conditions are not ideal for an echo canceller is an understatement. There could even be a time varying,

nonlinear condition in the path – a phone that has echo cancellation but just not enough echo cancellation.

An equipment vendor or carrier may reach the point where he/she is thinking about this difficult scenario only after getting past the “easier” problems. They may not start to recognize this problem until quite a bit of equipment is already fielded. But given the today’s lay of the land, the time will come. It’s better to design in the solution before fielding equipment rather than after.

The solution to this problem from an algorithmic standpoint is a not an easy one, but we have solved it at Adaptive Digital. Our recommendation is the use of one of our packet echo control algorithms that already has all the right pieces integrated.

4.2 Beamforming / AEC Combination

If you want an OK hands-free phone, you can use an acoustic echo canceller. But if you want to knock the socks off your competition, consider using a microphone array with both acoustic beamforming and acoustic echo cancellation. The beamformer will improve signal to noise ratio in the presence of background noise. It will point gain in the direction of the person in the room who is speaking, effectively bringing that person closer to the microphone array and reducing the relative level of room reverberation, which is a problem in larger rooms or when the person is not close to the microphone.

But what good is beamforming in a hands-free environment without acoustic echo cancellation? These are both very complex algorithms by themselves, but when they are put together, it’s tough to get it right. That’s why we have done it for you by making our beamformer and AEC easy to run together.