

**Line Echo Celler**  
**LEC-20 (RC-RLS)**  
**Specifications**  
**Rev. 3.0**

The data contained in the document is preliminary and is subject to change.

## TABLE OF CONTENTS

<b>1</b>	<b>SUMMARY.....</b>	<b>4</b>
<b>2</b>	<b>INTRODUCTION .....</b>	<b>5</b>
2.1	MAIN SOURCE OF ECHO .....	5
2.2	TRADITIONAL WAYS OF ECHO CONTROL: ATTENUATION AND SUPPRESSION.....	5
2.3	ECHO CANCELLATION.....	5
2.4	NATURE OF DOUBLE TALK.....	6
2.5	ECHO CANCELLER OBJECTIVES .....	6
2.6	DIFFERENCES BETWEEN NETWORK AND LINE EC FOR LAN-BASED SYSTEMS .....	7
<b>3</b>	<b>THEORY OF OPERATIONS.....</b>	<b>8</b>
3.1	ADF BLOCK.....	9
3.1.1	<i>Underlying Basic Operations.....</i>	<i>9</i>
3.1.2	<i>Reduced Complexity Recursive Least Square (RC-RLS) Method .....</i>	<i>9</i>
3.1.3	<i>Exploiting of Codec and Echo Path Properties.....</i>	<i>10</i>
3.1.4	<i>Convergence on Weak Signals .....</i>	<i>11</i>
3.1.5	<i>Step Size Tuning .....</i>	<i>11</i>
3.1.6	<i>RC-RLS Limitations.....</i>	<i>11</i>
3.1.7	<i>Echo Tail Length.....</i>	<i>11</i>
3.1.8	<i>Performance on Linear Codecs.....</i>	<i>11</i>
3.1.9	<i>Tone Auto Detection.....</i>	<i>12</i>
3.2	NLP BLOCK .....	12
3.2.1	<i>Post-Filtering.....</i>	<i>12</i>
3.2.2	<i>Comfort Noise .....</i>	<i>12</i>
3.3	CONTROL LOGIC BLOCK .....	12
3.4	IMPLEMENTATION ISSUES.....	13
<b>4</b>	<b>TYPICAL PERFORMANCE.....</b>	<b>14</b>
4.1	SINGLE TALK, LOW NOISE CONDITIONS. ....	15
4.1.1	<i>Adaptation on Preceding Background Noise. ....</i>	<i>16</i>
4.1.2	<i>Adaptation on the First Unvoiced Phoneme.....</i>	<i>17</i>
4.1.3	<i>Adaptation on the First Vowel.....</i>	<i>18</i>
4.1.4	<i>Adaptation on Another Vowel. ....</i>	<i>19</i>
4.1.5	<i>Further Behavior.....</i>	<i>20</i>
4.2	DOUBLE TALK .....	21
4.3	HIGH BACKGROUND NOISE, TANDEM, LOW-LEVEL DOUBLE TALK.....	22
4.3.1	<i>Selected Convergence Curves .....</i>	<i>23</i>
4.3.2	<i>Echo Path Estimation Errors. ....</i>	<i>24</i>
4.4	ECHO PATH CHANGE .....	24
4.4.1	<i>Sound of Inhaling .....</i>	<i>25</i>
4.4.2	<i>Greeting Words .....</i>	<i>26</i>
4.4.3	<i>Following Words .....</i>	<i>26</i>
4.4.4	<i>Echo Path Estimation Errors.....</i>	<i>27</i>
4.5	PERFORMANCE ON NOISE.....	27
4.6	PERFORMANCE ON LINEAR CODECS (ESTIMATED).....	27
<b>5</b>	<b>FORMAL TESTING.....</b>	<b>29</b>
<b>6</b>	<b>API .....</b>	<b>30</b>
6.1	DATA STRUCTURES.....	30
6.1.1	<i>Configuration.....</i>	<i>30</i>

---

6.1.2	<i>LEC Statistics</i> .....	30
6.2	IALG INTERFACE .....	31
6.3	VENDOR SPECIFIC API.....	31
6.3.1	<i>Initialization</i> .....	31
6.3.2	<i>Control</i> .....	31
6.3.3	<i>Process</i> .....	32
<b>7</b>	<b>APPENDICES</b> .....	<b>33</b>

## 1 Summary

The line echo canceller (LEC) is designed to provide the maximum attainable transparent voice quality for de-echoing of a PSTN or POTS connection in voice-over-LAN systems with internal delays, or on a codec end of a telecom switch. Its principal features are as follows:

- Very fast, typically 'the first phoneme' convergence due to the use of a regularized derivative of theoretically optimal Recursive Least Square (RLS) / Calman algorithm.
- Wide dynamic range (down to  $-65$  dBm); capable of converging on background noise even before parties start talking.
- Wiener-type post filtering to remove non-linear remnants of the echo (residual echo), to enhance full-duplex transparency during low-level double talk. That results in perceived double talk range beyond 40 dB on  $\mu$ /A-law codecs.
- Fast detection of echo path change.
- Recognition of tonal (any frequencies) or otherwise singular or narrow-band signals with poor eigenvalue spectrum, to prevent LEC misconvergence.
- Smooth transition between matching comfort noise and real background noise, to result in perceptually insignificant noise contrast.
- Achieves 33-35 dB of ERLE if  $\mu$ /A-law codec are employed.
- Achieves up to 55 dB ERLE, if linear codecs are employed.
- Active echo tail capacity of 20 ms.
- Packet-based processing with 5 ms frame size.
- Fully utilizes C55x architecture and instruction set advantages.
- Reasonable MIPS (7.5 MHz), object (1022W), scratch (282W) and program memory (9.2KB) requirements.
- Bulk delay should be either 0 or known: the LEC should be attached to the codec interface. This version of LEC is not a network echo canceller to be placed on a PSTN's backbone.
- No tone disabler capability is currently provided due to the limited application target (not intended to work with voice-band modems or facsimile machine).
- Packet Loss Concealment will be incorporated in subsequent releases.
- Quality Voice Activity Detection will be incorporated in subsequent releases.
- Automatic Gain Control will be incorporated in subsequent releases.
- Designed as a system component rather than a stand-alone algorithm; provides statistics and is capable of non-intrusive monitoring.

## 2 Introduction

### 2.1 Main Source of Echo

Echo is unavoidable in today's telephony networks. The reason is purely historic: the first phones had 2-wire connections to central offices (COs), and they employed devices called hybrid to de-couple a microphone and a loudspeaker. This type of 2-wire connection is still used in the most local loops.

Hybrid is essentially a balanced transformer, and it may not be made ideal. Any mismatch leads to signal reflections, perceived as an echo. The human binaural auditory perception system is quite capable of effectively masking many types of echo in natural environments. For example, two people can talk in a big room with plain parallel walls (which produce long-lasting high-level multi-pass reflections), and still understand each other perfectly because the ears and brain can filter out such echoes very effectively. Unfortunately, this is not the case with the echoes produced by a telephony network. These echoes are different. They are strong, mono, localized in time domain and quite unnatural: the human ear cannot completely filter them out. There is an echo from the local hybrid, coming back with negligible delay. This is a so-called side-tone, which people actually perceive positively as an indication that they are on-line. Additionally, there are delayed echoes from the far end's hybrid(s). Such delayed echoes can be very annoying, and the annoyance increases with round-trip delay and the echo's increased intensity. The echo is most audible on transitions between phonemes: an echo of a vowel becomes more apparent when the vowel is followed by silence or a sound with different spectral properties.

### 2.2 Traditional Ways of Echo Control: Attenuation and Suppression

The first approaches of dealing with far-end echo were the insertion of attenuation and echo suppression. If the delays are moderate, less than 5-10 ms, then a reasonable amount of attenuation (3-9 dB) helps to mask the far-end echo by the local side-tone. This attenuation applies only once to the one-way voice, but twice to the echo, going forth and back. The amount of attenuation that is required for echo masking becomes objectionable if delays increase beyond 20ms. Such delays normally appear on long-distance calls. For example, a coast-to-coast call, routed by the circuit-switching PSTN (where 200 km of distance roughly corresponds to 1 ms of one-way delay), has at least 20-25 ms of one-way delay. So-called echo suppressors were used in these cases.

Echo suppressors allowed only one direction to be enabled at a time, and transmission in the opposite direction was disabled. Echo suppressors blocked round-trip echo path, but effectively enforced nearly half-duplex mode of communication on long distance calls and caused many other problems (see ITU-T G.168). People could not speak at the same time (so-called double talk condition), which is the way many people naturally communicate. Some people worked out a habit to communicate differently on local and long distance calls. However, most people were still assuming conditions of a face-to-face meeting where everything that was said was heard. This resulted in many misunderstandings, communication losses, and inconvenience.

### 2.3 Echo Cancellation

The technology progressed, and first echo cancellers (EC) appeared. They were designed to allow both parties to talk and be heard at the same time, to some extent.

In the beginning, ECs were extremely expensive and very far from perfect. As with any new technology, ECs improved over time and became much more affordable. But even today, state-of-the-art commercial ECs are far from ideal. Anybody spending significant time making long distance and overseas calls will notice the occasional echo and severe voice clippings. If you are sufficiently familiar with EC technology and you know its potential weak points, it will not take you long to force almost any EC in today's network to operate inappropriately just by talking at certain times with certain intensity.

The quality of an EC reveals itself during double talk periods of a call (if no double talk happens, a properly designed EC is not different from a traditional echo suppressor), and it depends on many

circumstances. In certain conditions, even the best EC-equipped circuits produce either half-duplex connection (no double talk allowed) or generate an undesirable echo.

## **2.4 Nature of Double Talk**

An ideal conversation between two people with perfect emotional self-control is unlikely to have any double talk periods. One speaks for a while, and then pauses, signaling the other that it is now his or her turn to talk. The other party keeps silent and patiently waits for the pause as an invitation to start talking, never interrupting. If such a (very uncommon and unrealistic) scenario were always the case, there would be no need for voice echo cancellers because echo suppressors would suffice.

Many people acknowledge understanding by an audio equivalent of nodding by saying something like ‘yeah...ok...hmm... well...’, in a soft voice. Many people misjudge pause length and step in while another party is still talking. Many people do not have perfect control over emotions and do interrupt. Conversation patterns vary depending on age, status, gender, regional specifics, and so on. The probability of double talk increases for conference calls, when three or more people need to determine whom is to step in after a pause. Large end-to-end delays also add to double talk intensity because they alter the perception of the natural flow of the conversation. We are not yet in the video-telephony era, and many signals that people successfully use in face-to-face meetings cannot be used if the media is audio-only.

Most of the double talk conditions can be classified as low-level double talk, or as fast interleaving of activities between parties, when the start of the far-end activity overlaps with the end of the last near-party’s word, and vice versa. The voice levels may change very quickly. In many languages, the most important parts of information are passed through short, low energy, unvoiced sounds while longer, higher energy, linking vowels pass much less information. If a far-end party pronounces “4 16” in a soft voice while near-end party is almost yelling, ECs should not clip these sounds, neither completely (as an echo suppressor would do), nor partially, modifying the signal by removing short low-energy fricatives and producing a distorted “or 60”. These sounds may bear a quite different or even opposite meaning than intended.

The main goal of a voice echo canceller is to help keep most real-life conversations running smoothly and perceptually undistorted despite imperfect nature of telephony networks.

## **2.5 Echo Canceller Objectives**

Generally, the deeper the double-talk range of an EC, the better it is. An ideal LEC preserves even the sound of breathing on the far-end while near-end speaker is talking, maintaining the illusion of full transparency. A good EC passes the signal of a far-end party, even if that party speaks with a much softer voice than the near-end party, producing no objectionable voice clippings, nor exposing the near-end party to echo remnants, in wide range of conditions, which occur in the PSTN. Such conditions include the following:

- Various (and possibly changing) echo paths, various return echo levels, and so on.
- Various voice levels, different speaking patterns, different languages, and so on.
- New parties stepping into the conversation (such as in a conference call), with varied voice spectrum and pitch (for example, a change from a deep male bass to a soft, high-pitched, female voice).
- Changing of volume on any side, muting, switching to hands-free and back (more noise, lower level of the speech), and so on.
- Various (and possibly changing) background noise levels, which can include impulse noise and music on hold, and so on.
- Correct processing of various tones like DTMF and CP, starting from the beginning of the conversation.

Interworking (tandem conditions) with another PSTN echo canceller, acoustic echo controller (even with imperfect echo reduction), on both ends.

It is difficult to represent the performance of an EC in numbers. Internal measures like echo return loss enhancement (ERLE) or convergence speed (which are explained later in this document) on noise-like signals may be identical for both poor and high quality echo cancellers. Properly organized subjective testing reveals that most EC qualities, when mere statements of conformance to standards, may be misleading. A good EC usually has high marks on formal tests. The opposite statement is not true—an EC with high ERLE measures may sound awful.

## **2.6 Differences between Network and Line EC for LAN-Based Systems**

Fortunately for the network EC, general public expectations of long distance call quality are still relatively low. Mediocre systems find customers, if the price is appropriate. Low-bit-rate channel coding by a CELP vocoder additionally suppresses or removes echo remnants thus masking the EC's imperfections.

In contrast, public expectations of local call quality are very high. People expect fully transparent sound, no echo, no distortions, and no clippings. Thus, the line echo canceller (LEC) for packet-based voice over LAN systems without voice compression is different from a network EC, in the following ways:

- Network EC is not exposed to a tonal exchange at the beginning of outgoing calls and the LEC is. As a result, some network EC requirements shall be modified.
- Voice band data modem and fax machines are not usually connected through LAN infrastructure, but use dedicated connections. Therefore, the LEC does not need to comply with corresponding interoperability requirements.
- LAN inserts significant amount of one-way bulk delay (node to node, and between TDM interfaces) due to packetization and de-jittering activities, typically between 20 and 50 ms. Such delay is similar to that present in overseas calls.
- LEC can be explicitly reset at the beginning or end of every call, so the importance of fast echo path change detection is much lower than for a network EC.
- LEC should not start a call with non-linear processor NLP disabled because user-audible call flow differs from a regular local call.
- LAN bandwidth is sufficient to avoid the use of voice compression.
- LAN-based PBXs frequently use high quality feature phones with volume controls up to +15–+20 dB. When people adjust the volume, they also enhance the audibility of the LEC's undesirable byproducts. Any artifacts produced by the LEC become highly audible.
- LEC should maximize the double talk range while the importance of other objectives may be partly lowered.
- A good network EC does not necessarily make a good LEC, and vice versa.

Overall voice quality of LAN-based PBX-like systems depends mostly on the quality of two signal processing algorithms: jitter buffer and LEC. The LEC, working on the PSTN or POTS interface, should be nearly perfect. If a call is known to parties to be essentially local, then any echo-connected effects, easily forgiven on a long-distance call, will be reported as a significant problem. Eventually, problems with the LEC may lead to the situation in which the system is returned to the vendor.

### 3 Theory of Operations

This LEC-20 is optimized to deal with the specific problem of echo cancellation for packetized-voice-over-LAN systems, with significant internal delays, to create and maintain the user's perception of an undistorted, transparent, full-duplex PSTN connection. It accumulates years of experience in LEC development, real-life field deployments, troubleshooting, and maintenance. Its performance was verified using a large base of recordings of real-life trouble situations with sudden echo path changes, various double talk conditions, tones, overloads, non-linear echoes—altogether, representing the worst-case scenarios that nevertheless happen from time to time.

Voice is very different from noise signals traditionally used by simplified EC test procedures such as G.165<sup>1</sup> or G.168<sup>2</sup>, and the results of those tests are hardly representative. The following sections provide further information clarifying the details of the LEC-20's internal structure. That should help set expectations appropriately and to avoid misconceptions.

The LEC consists of three major blocks:

- adaptive filter (ADF)
- non-linear processor (NLP)
- control logic

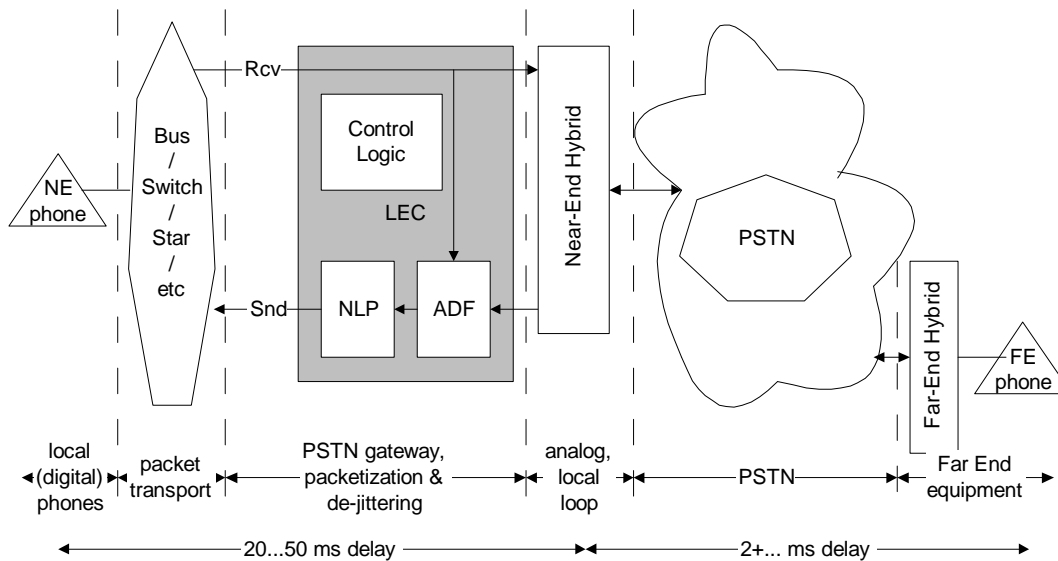


Figure 3.1. LEC, its major blocks, and its place in the infrastructure.

Some echo cancellers mix the Rcv path signal with white noise or pseudo-random maximum length sequences, continuously or temporarily, to speed up initial convergence and to distinguish between double talk and the echo path change. This leads to achieving high scores on formal tests but this noise is noticeable and annoying to listeners. This LEC does not use such methods because that would contradict the main LEC objective of maximizing perceptual quality.

<sup>1</sup> ITU-T G.165 (1993), Echo Cancellers.

<sup>2</sup> ITU-T G.168 (2002), Digital Network Echo Cancellers.

## 3.1 ADF Block

### 3.1.1 Underlying Basic Operations

The essence of the ADF operations is simple:

- ADF assumes that the echo path is somewhat linear.
- ADF remembers what was the reaction of the echo path on the signals previously sent to the far-end.
- If the current signal can be (partially or fully) represented by a linear combination of previously used signals, then ADF may predict the echo path reaction on this signal. This reaction is subtracted from the real input from the far-end, effectively cancelling the echo. The remaining signal shall be far-end speech.
- The result of most recent processing is then used for further adapting (continuous learning).
- Learning results in the building of the most appropriate model of the echo path.

This is a classical adaptive approach, well studied by many scientists. There are many adaptive algorithms, ranging from simple Normalized Least Mean Square (NLMS) to theoretically optimal Recursive Least Square (RLS) method. All are intended to solve a system of linear equations (but in a recursive approach):

$$h_k = (X_k^T W_k^{-1} X_k)^{-1} X_k^T W_k^{-1} y_k;$$

where:

- $h_k$  is the echo path model,
- $W_k$  is a correlation matrix of noise (generally, unknown)
- $y_k$  is a vector of the signal received from codec what is a mix of echo and a far-end speech
- and  $X_k$  is a matrix of signal sent towards the codec, which is, human voice or tones

### 3.1.2 Reduced Complexity Recursive Least Square (RC-RLS) Method

ADF employs a regularized derivative of true RLS method with reduced complexity (RC-RLS), coming as close to the original RLS as possible, while keeping it stable in real-life conditions on a 16-bit fixed point DSP. The reasons for RC-RLS introduction are given in this section.

NLMS-like algorithms share very similar performance with true RLS, if the excitation signal is white noise or zero-auto-correlated in other meanings (like delta-function), what is the ideal input signal. Both RLS and NLMS degrade if the eigen-value spectrum of the Fisher matrix ( $X_k^T W_k^{-1} X_k$ ) of the input becomes uneven. The NLMS degradation (absolute or relative to RLS) is minimal if the spectrum of the input signal has a gradual gentle roll-off, like CSS of G.168.

Human voice is essentially highly auto-correlated due to the complex nature of vocal tract excitation. As the result, the Fisher matrix ( $X_k^T W_k^{-1} X_k$ ) has a poor spectrum of eigenvalues, both in the short-term and the long-term. The problem of solving close-to-singular systems of linear equations is well known in the tensor and functional analysis, and it is one of the central problems in the theory of numerical recipes as a matrix pseudo-inversion problem. Simple methods like gradient search do not converge with acceptable quality and speed. NLMS-like algorithms degrade very quickly if the input is periodic and/or its spectrum has strong formant structure, while RLS is still capable of converging (further notes on the analysis of NLMS performance on voice-like signals are in Appendix B). Double talk range becomes low for voice-driven NLMS-based ECs. As the LEC is required to exhibit nearly perfect performance on a real-life voice signal, more advanced adaptive algorithms than NLMS are required.

Optimal algorithms are more sensitive to the violations of the assumptions under the classical adaptive approach. Voice echo cancellation represents a case with serious deviations from these assumptions:

- Echo path is not exactly linear, echo may not be completely cancelled
- Distortions introduced by codecs result in so-called 'noise on input', when exact signal entering the echo path is unknown
- Echo path is not stationary, in theoretically peculiar ways

- Voice represents a signal with poor eigen-value spectrum
- Properties of far-end signal are unknown a priori
- Exact computations are impossible, especially on a commercial affordable 16-bit DSP. That alters the meaning of operations and impacts performance

There are very many theoretical articles<sup>3,4,5</sup> and books<sup>6,7</sup> providing detailed performance analysis, but all under idealized simplified conditions. There is a very strong assumption of linearity under either RLS or Calman filter equations, coming from the underlying Gramm – Schmidt orthogonalization algorithm. This algorithm does not work if this assumption is broken. Any RLS-based EC will inevitably explode, as far as signal decomposition progresses. The echo path non-linearity and the voice spectrum's singularities create together a problem well beyond an impact of each separate reason. That was the major reason of reworking RLS into a more robust algorithm, referred here as RC-RLS, which takes into account the deviations mentioned above.

The RC-RLS is not the only method suitable for the voice echo cancellation applications. There are other methods, all of them superior to NLMS, as affine projection, transform domain algorithm, etc. Most of them concentrate on fixed or adaptive pre-whitening of the current voice signal. Many of them have been tested in the process of the ADF development, but have been found to be inferior to the RLS derivative for this particular application, for various reasons. A detailed explanation of these reasons goes beyond the scope of this document.

Even a theoretically optimal RLS cannot recover echo path curve if the input signal does not cover the entire spectrum. If echo path response on certain signals is not observed, RLS makes no assumptions. Fortunately, there are certain properties of the codecs (because they must comply with ITU G.712<sup>8</sup>) and telephony circuits that determine how the echo path will and will not look like.

### 3.1.3 Exploiting of Codec and Echo Path Properties

The properties of codecs' receiving (Rx) and transmitting (Tx) filters have major impact on the echo tail shape. The frequency responses of these filters have the narrowest transient bands (below 300 Hz and above 3300 Hz), compared to other devices along the circuit. They determine the echo tail disperse region, and otherwise are flat in their pass-band. Analog circuitry and local loops (which are shorter than 10 km  $\approx$  40 $\mu$ s) may introduce gradual roll-off or emphasis. Two or more reflections with delay between them result in a frequency beat with spacing inversely proportional to the delay and the amplitude, depending on the difference between reflection magnitudes. Generally, the frequency response of the echo path is reasonably smooth. Any echo path tail shares common properties and is devoid of impossible ways of existence.

There are many ways in which this a priori information can be used, and RC-RLS exploits them to enhance its convergence. Moreover, RC-RLS assesses the quality of adaptation by estimating the Mean Square Error (MSE). Even if the excitation signal is purely periodic (like a vowel with wide short-term spectrum), and therefore the corresponding Fisher matrix ( $X_k^T W_k^{-1} X_k$ ) is singular, the echo path can still be estimated by interpolation techniques if the pitch is lower than the frequency beat spacing.

See Appendix A for more information on the echo path properties.

---

<sup>3</sup> J.M. Cioffi, T. Kailath. Fast, Recursive-Least-Squares Transversal Filters for Adaptive Filtering, IEEE Trans. Vol. ASSP 32, No.2, April 1984.

<sup>4</sup> A. Houacine. Regularized Fast Recursive Least Square Algorithms for Adaptive Filtering, IEEE trans on Signal Processing, Vol. 39, No.4. April 1991.

<sup>5</sup> M. Zoltowski. Why do optimal forgetting RLSs exhibit long term divergence and how can this be avoided?, Elsevier Signal Processing 68 (1998).

<sup>6</sup> P. M. Clarkson. 'Optimal and adaptive signal processing', CRC Press, 1993

<sup>7</sup> N. Kalouptsidis, S. Theodoridis, 'Adaptive system identification and signal processing algorithms, Prentice Hall, 1993.

<sup>8</sup> ITU-T G.712 (1996), Transmission performance characteristics of pulse code modulation channels.

### 3.1.4 Convergence on Weak Signals

The signals of low-level fricatives, consonants, breathing, and background noise have a much smoother spectrum than the spectrum of strong periodic vowels. RC-RLS overcomes 16-bit arithmetic and the DSP instruction set's limitations and extends the dynamic range of adaptation down to the -55...-65 dBm range. Adaptation quality and speed are affected by the quantization of low-level signals and their echoes in the codecs, as well as by the Snd noise level. Nevertheless, ADF can adapt to background noise and achieve a range of approximately 15 to 20 dB of potential echo cancellation in 200 to 500 ms, in a typical call scenario, before first word even is pronounced.

### 3.1.5 Step Size Tuning

Another strong assumption is that adaptive algorithms know the statistical properties of the disturbing signal (far-end, send direction, noise). This assumption is false, and the classical algorithms of optimal step-size tuning do not apply. RC-RLS uses an additional layer of adaptation to allow adequate step-size tuning, and unties step-size tuning from a binary decision of double talk detector, which controls onlyNLP.

### 3.1.6 RC-RLS Limitations

A standard  $\mu$ /A-law codec introduces very specific non-linear distortions, and they limit echo return loss enhancement (ERLE) of any echo canceller to a range of approximately 32 to 35 dB, even if there is a single  $\mu$ -law codec in the echo path.

ADF converges very quickly on most, but not all, voice excitation signals. The RC-RLS convergence speed and the corresponding double-talk range vary depending on the spectrum and amplitudes of voice and noise. RC-RLS also loses convergence quality if the echo path non-linearity significantly breaches requirements of G.712 (see Appendix A).

Further details on RC-RLS structure and algorithmic base are beyond the scope of this document.

### 3.1.7 Echo Tail Length

The required echo tail capacity for network and line echo cancellers has been a subject of multiple debates. Currently, PSTN is accomplishing another spiral of evolution. Digital Circuit Multiplication Equipment with high-quality voice compression has recently become much more affordable and widespread, and long-distance calls have undergone a sharp cost decrease. Network EC stations have become much cheaper, and they tend to move closer to the customer premise equipment (CPE) every year

The requirements of a network EC and a LEC are very different from the point of view of echo tail length, and the LEC does not have to be a copy of the G.168 network EC. This LEC provides 20 ms of active echo tail, and this has been found sufficient and adequate for all cases previously observed in the field. The LEC tends to switch into half-duplex mode if its echo tail capacity is exceeded. Longer echo path capacity may be provided in following releases, depending on customer requirements.

There are still regions in the world where echo tails are longer than 20ms. These regions are diminishing. Besides having multiple reflections (up to 10), the circuits in these regions suffer from high noise, low signal levels, and various types of non-linearity. These circuits are frequently compressed by low-delay DCME (ADPCM or LD-CELP). Mere EC with 48/64/128ms echo tail will unlikely sound significantly better than an echo suppressor under such conditions. If you desire to extend the full-duplex coverage to these regions, you will require extensive voice enhancement solution (instead of just EC), equipped with far-end (both linear and non-linear) echo control capability, automatic gain control, noise subtraction and other similar valuable features.

### 3.1.8 Performance on Linear Codecs

The LEC may improve double-talk range by 10 to 25 dB if linear codecs are used. Traditional PSTN infrastructure uses  $\mu$ /A-law codecs, but good 14/16 linear codecs with low non-linear distortions are

currently very affordable. They are widely used by voice-band modems. If a voice-over-LAN system uses such codecs, the near-end echo (from the adjacent hybrid) is reproduced in the linear domain. The far-end codec is behind a  $\mu$ /A-law codec-based CO and its echo response is still corrupted. PSTN and local loops usually insert about 6 dB of one-way attenuation on local calls<sup>9</sup>. This typically limits the enhancements by 7 to 12 dB.

An exploitation of the benefits associated with linear codecs requires that the dispersion time of an echo tail is longer than 8 ms. If the chosen codec does not comply with frequency response requirements of G.712, external DSP software filters must be provided. Insufficient echo tail capacity often results in excessive echo in the frequency ranges closest to the codec filter's transient bands. RC-RLS uses block-floating-point representation of the echo path to preserve precision.

### 3.1.9 Tone Auto Detection

ADF cannot reliably converge on single or double tones because their spectrum is too singular. Instead, it studies the distribution of eigen-values and disables adaptation if the spectrum does not look right.

If DTMF tones are partially clipped because they have been amplified, they may not be recognized as pure sine waves. LEC provides threshold settings via control interface to accommodate that situation.

## 3.2 NLP Block

Usually, NLP removes the echo remnants (residual echo) during single talk periods if they are not masked by a double talk signal, or if they are not buried under background noise. When double talk is about to be detected, NLP starts to open gates to the far-end signal, and then residual echo may be heard, depending on delays, levels, and spectrums of the signals.

Echo cannot be cancelled perfectly by an ADF based on a linear model (the best fit for the current signal), owing to the echo path non-linearity. The residual echo for  $\mu$ -law codecs in the case of converged ADF represents a non-linear transform of the transmitted signal, and its spectrum is related to the spectrum of original signal.

### 3.2.1 Post-Filtering

The NLP is enhanced to exploit this property of residual echoes. The residual echo can be filtered so that it is masked better by the far end signal, and only the perceptible part is clipped. This post-processing is similar to the Wiener filter. It weights the estimation of non-linear residual echo and incoming near-end signal, taking into account perceptual masking effects. The result of the filtering often sounds as if high frequencies are attenuated while the far-end signal is still intelligible. This operation improves the sound especially in the cases of low-level double talk, when it is too easy for NLP to clip out an initial weak phoneme.

### 3.2.2 Comfort Noise

The NLP clips or removes audible remnants of the echo, and inserts comfort noise instead. Although many efforts have been spent on matching the artificial comfort noise to the real background noise and smoothing transitions, there still may be cases when switching is audible. There may be real background noise with an unusual spectrum, or very soft but still audible sounds of music mixed with background noise. The NLP may remove such music together with the residual echo, and insert an artificial comfort noise not matching the melody.

## 3.3 Control Logic Block

The control logic is required to bind the algorithmic blocks and sub-blocks of LEC together and guarantee their smooth interactions. The control logic block is responsible for:

- Detecting channel activity and double talk conditions

---

<sup>9</sup> BellCore SR-TSV-002476 (December 1992) quotes surveyed median loss on local loops as 3.5 dB (on 1000 Hz).

- Switching from half-duplex into full-duplex mode (and back)
- Controlling over adaptation modes
- Setting appropriate NLP mode
- Supporting control interface and applying required configurations
- Checking and maintaining LEC integrity
- Collecting and reporting statistics of LEC performance

Further details are beyond the scope of this document.

### **3.4 Implementation Issues**

Adaptive filtering is a classic example of a MIPS and memory demanding application. It is also an example of a non-robust algorithm: to achieve a certain precision on output, LEC must use much higher internal precision. The discussion of underlying reasons is beyond the scope of this document.

The algorithm is heavy in term of MIPS usage. Even on the very powerful C55x core with high degree of parallelism, even using linear code (to avoid call or branch overheads) and fully assembled core routines (60 % of entire code size); the LEC consumes about 5.5 MHz per channel just to run adaptation. About 2 additional MHz are required to run the EC control logic, signal analysis, NLP, etc. A typical LAN-based PBX has 4 to 12 PSTN connections, so there is not much saving in the bulk mode, when many instances of the LEC negotiate which is to run an adaptation at a given moment.

There is not much sense in improving MIPS usage on account of the LEC quality because the hardware associated with LEC represents only a tiny fraction of the system cost. Any simplifications undertaken to lower MIPS and memory usage will likely affect EC performance, and this is undesirable in voice-over-LAN applications.

## 4 Typical Performance

ITU-T P.50<sup>10</sup> recommends simulating real speech signals by using models with varying pitch, different for male and female speakers, unvoiced sounds 4 times shorter and 17.5 dB lower than the voiced sounds, and changing short-term spectrum.

The CSS signal, used by G.168, is overly simplistic and does not conform to ITU-T P.50, neither in variability of pitch and spectrum, nor in relations between levels and duration of voiced and unvoiced sounds. If CSS were a realistic speech representation, there would be no need in EC methods other than NLMS because this signal is very easy to adapt on. However, there are no simple artificial test suites exhaustively covering the LEC performance in real conditions. The LEC-20 was tested versus a large base of recordings of real-life trouble situations with sudden echo path changes, various double talk conditions, tones, overloads, non-linear echoes—altogether, representing the worst-case scenarios that nevertheless happen from time to time.

The following test cases provide readers with detailed systematic performance analysis for typical call scenarios. Performance of LEC is reported upon completion of every stage of call:

- In the terms of energy (Rcv, Snd, ERLE, MSE, etc),
- In spectral terms for the signal, its echo, and its residual error (which determine audibility and perceptual quality).
- The spectral analysis of echo path estimation error is also provided. It determines how LEC behaves at the start of next phoneme, which spectrum is given in the description of next stage.

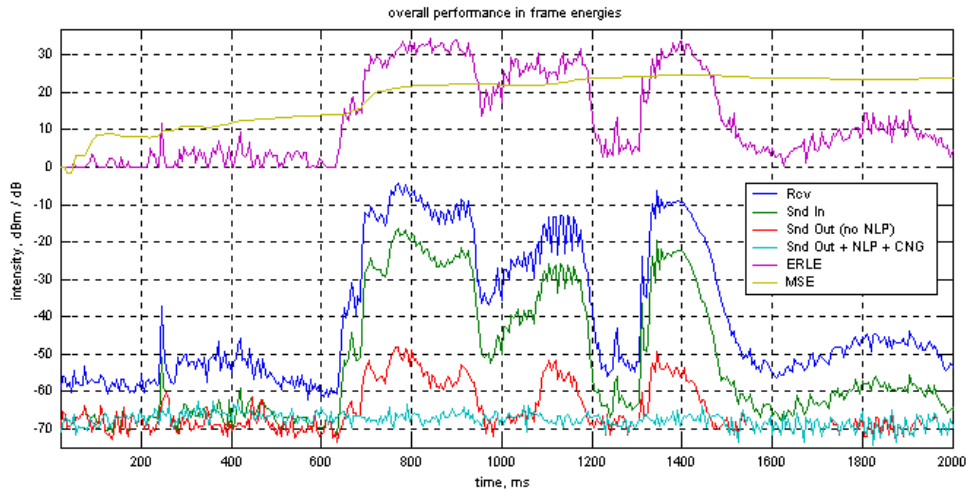
NOTE: The examples given below do not represent the best cases of LEC-20 performance.

---

<sup>10</sup> ITU-T P.50 Artificial Voices (1993)

### 4.1 Single Talk, Low Noise Conditions.

The following pictures illustrate LEC performance in a good case of a call when there is neither double talk nor significant background noise in the Send path. Codecs are  $\mu$ -law. Initial phase of tonal exchange before connection is set is omitted simplicity sake. The speaker (male, English language) first inhaled and then said the sentence: ‘Hello, is Mike there?’ followed by another inhale (Figure 4.1). The conversation continued, and after it was over, white noise was sent to the echo path. All signals were recorded, thus allowing detailed analysis of LEC performance.



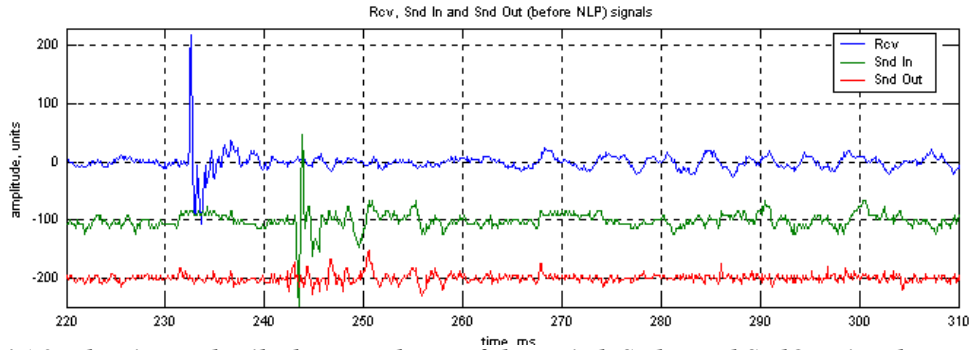
**Figure 4.1.1.** The MSE is as it is self-accessed by LEC-20. ERLE was measured using 5 ms frames.

The sentence occupies time from 630 ms until 1600 ms (there are no pauses in the actual sound). The holes in the energy plot correspond to the sounds of consonants and fricative.

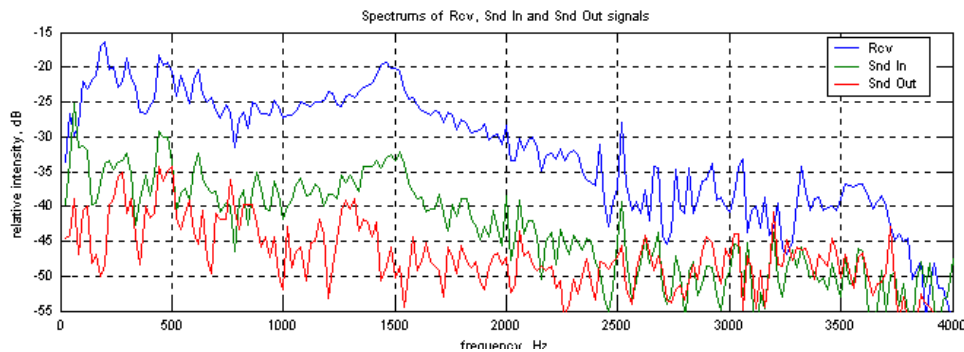
You can see that LEC successfully pre-adapts on the sound of inhaling and the occasional click at 250ms. The first [h] sound is almost immediately cancelled down to the level of background noise (-65...-70dBm). The following stronger [e] vowel is cancelled by almost 30 dB from the very start. NLP is switching in immediately, as residual echo may be audible. The level of comfort noise is somehow excessive at the beginning, but it will adjust later (in 1.5 s ... 2s). The following sound [ai] has lower pitch (what is noticeable by the fluctuating packet energy, 1100ms to 1200 ms); the ERLE is above 25 dB for the duration of the sound. The self-accessed MSE measure is more conservative: it achieves 14 dB during initial pre-adaptation and rises to about 25 dB after the first sentence is over.

### 4.1.1 Adaptation on Preceding Background Noise.

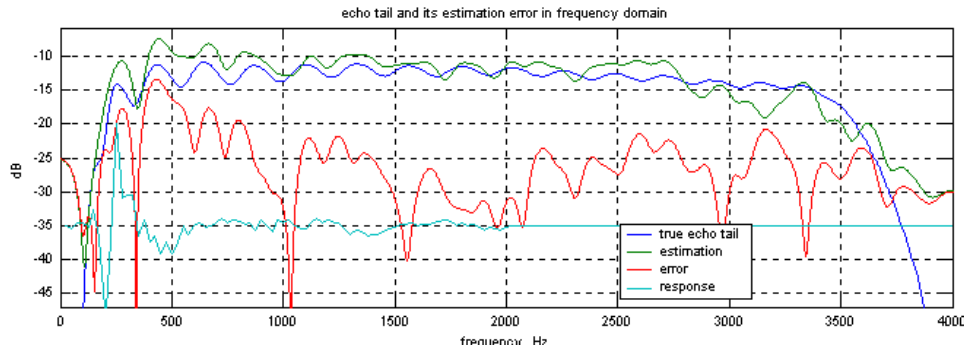
The first phase is inhaling with an occasional click; lasting for the first 600 ms. The signals have average intensity of about  $-55$  dBm with heavy quantization. Low frequency hum (Rcv line noise) partly is partly filtered by LEC's input band pass filter.



**Figure 4.1.2.** The picture details the central part of the period. SndIn and SndOut signals are amplified by 12 dB and shifted down for clarity. Note heavy m-law quantization of SndIn signal. Note that even the first short click can be cancelled<sup>11</sup>. This click is followed by low-frequency signal, which is not cancelled as well.



**Figure 4.1.3.** The spectrums of Rcv, Snd In and Snd Out (before NLP) for the first 0.5 s are similar. There are some improvements for mid-frequency components, but it is accompanied by degradation above 2500 Hz.



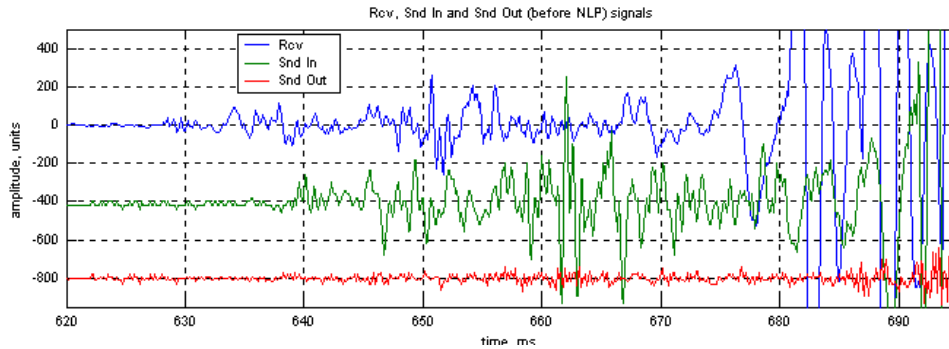
**Figure 4.1.4.** The actual estimation of echo tail in time domain, shifted down and amplified, is plotted in cyan. LEC avoids modifying echo tail in the regions where the variance of estimation is compatible with estimation itself<sup>12</sup>. Pre-adapted LEC is now capable of achieving about 15 dB of ERLE in the frequency range between 800Hz and 2500Hz, where human ear is most sensitive.

<sup>11</sup> Actually, short  $\delta$ -function-like signals with flat and wide spectrum are the very easy for adaptive filters to adapt on if signal normalization precautions are observed in a fixed-point EC implementation.

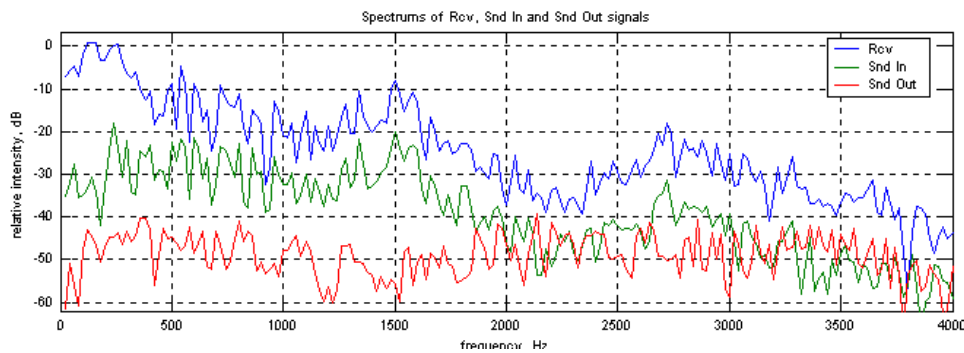
<sup>12</sup> This is a basic principle of biased estimation.

### 4.1.2 Adaptation on the First Unvoiced Phoneme.

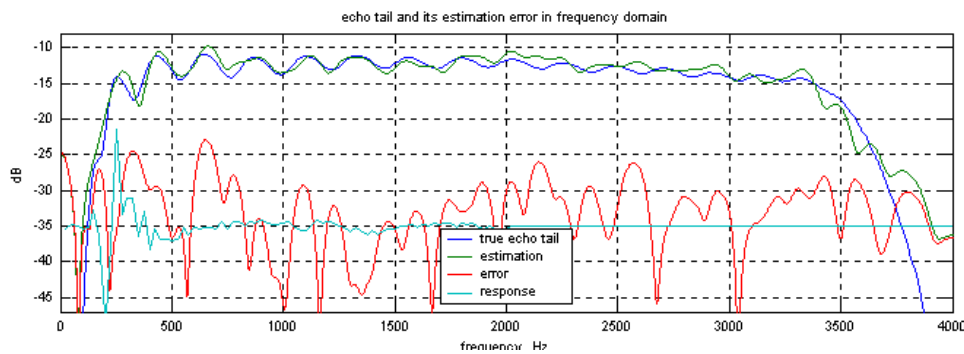
The sounds of inhaling followed by ‘Hi...’ or ‘Hello...’ are probably the most common sounds in the beginning of calls in North America (and in many other regions). Thus the next phase is a very important phoneme [h] of the ‘Hello’ word. It is a fricative, noise-like sound with duration about 50 ms and energy – 35...-45 dBm. Growing signal of [e] sound preempts at 675ms.



**Figure 4.1.5. Snd In and Snd Out signals are amplified by 20 dB and shifted 400 and 800 units down. There is slight DC bias in Snd In signal. The residual echo stays on the background noise level for the most of [h] sound duration.**



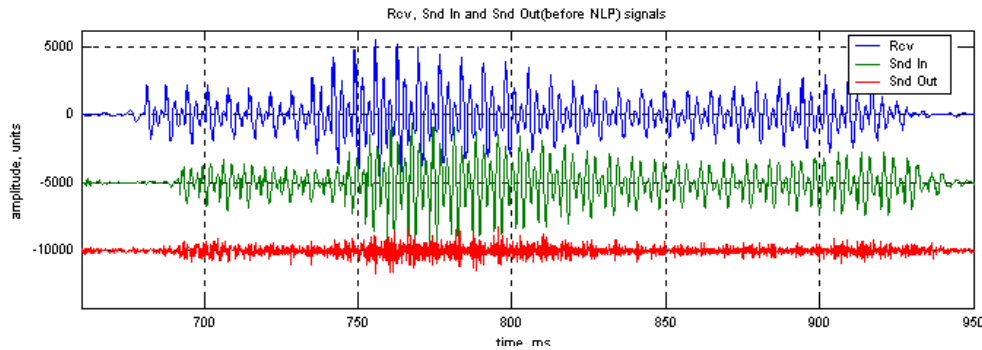
**Figure 4.1.6. The effect of pre-adapting is visible. The most of the echo's spectral content is cancelled by 15...20 dB, except for areas with low spectral density (2000...2500 Hz).**



**Figure 4.1.7. LEC utilizes 50 ms of low energy (-40 dBm) sound [h] to achieve the cancellation quality of about 15...20 dB, spread almost evenly across full frequency range.**

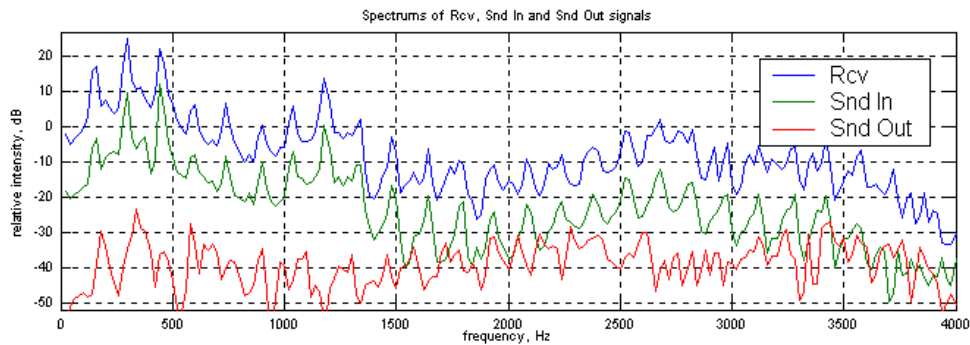
### 4.1.3 Adaptation on the First Vowel.

Vowels are periodic signal with relatively stable pitch (period), depending on emotional context and vocal abilities of the speaker, high amplitude, slow rising and falling edges, and relatively long duration (in order of 100s ms), except for the cases of frequently used words. The sound below is actually 3 vowels (and [L] between them) pronounced without a pause: *'h/-[e\_llo\_i]-/s'*.

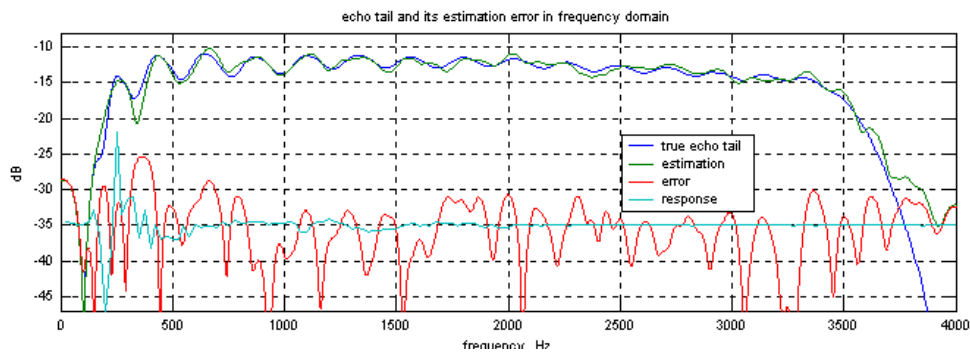


**Figure 4.1.8.** *Snd In* signal is amplified by 12 dB (~ERL), and *Snd Out* signal is amplified by 32 dB. The ERLE was about 26...28 dB in the beginning of the opening vowel [e], and it risen to 32...34 dB by the end of the sound. Note that *Snd Out* signal does not look as periodic as *Rcv* or *Snd In* signals do.

Let's analyze the LEC reaction on the first vowel [e], lasting from 680 to 740 ms.

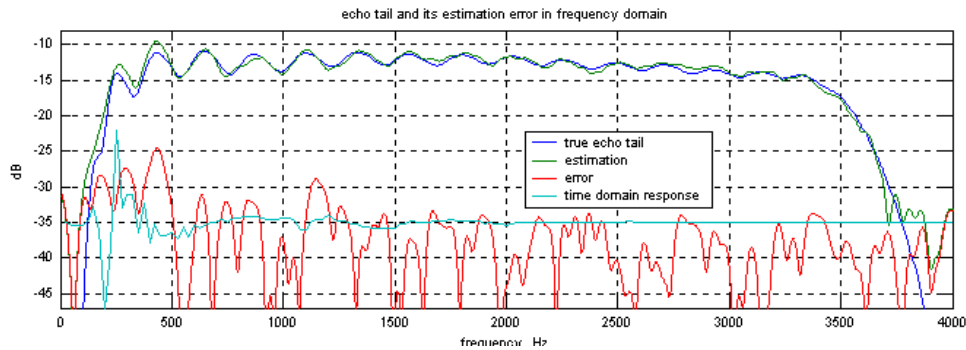


**Figure 4.1.9.** The spectrum of *Snd Out* signal does not repeat the shape of *Snd In* signal spectrum. High-energy low-frequency components are cancelled by 30...40 dB. Lower energy components (above 1500 Hz) are cancelled as well, but with lower quality.



**Figure 4.1.10.** The spectrum of echo path estimation error goes down and becomes more even, although the quality improvements are less drastic than after [h]. The frequency response of the error is about 20 dB below echo path response for signals above 500 Hz.

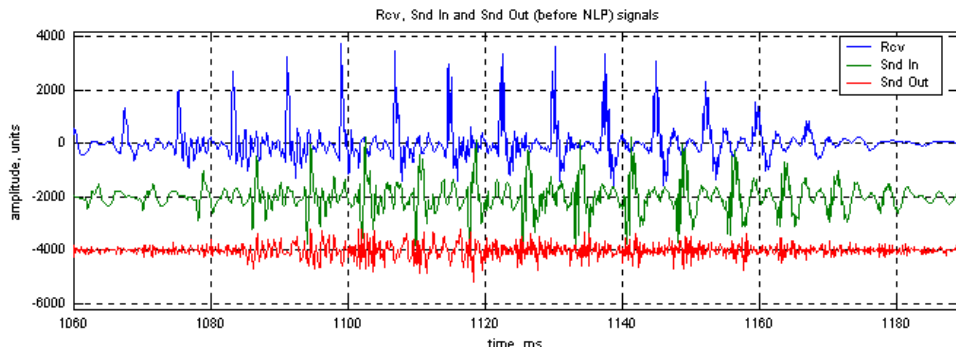
The spectrum of the echo path estimation error after entire voiced sound of Figure 4.8 finishes is illustrated below.



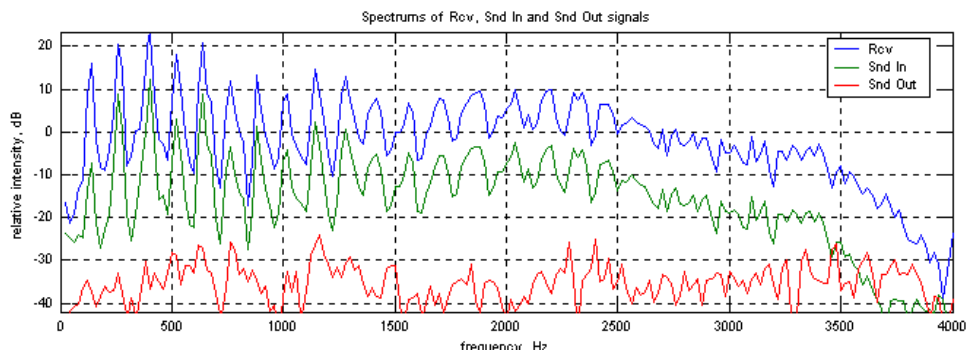
**Figure 4.1.11.** Now the error spectrum is almost even, about 25 dB below true echo path response for signals above 500 Hz. This number is connected to MSE and very closely corresponds to true double talk resolution. To achieve high double talk resolution on low frequencies, the echo tail should be studied for the length about 12...15 ms due to the slow decay of the oscillations corresponding to the sharp high-pass filter incorporated in G.712 complying codecs.

At the end of this compound signal, LEC starts to consider itself as preliminarily adapted and may detect double talk conditions. The following convergence is aimed to enhance the precision further and stretch double talk range as deeply as possible for all signals that may happen.

#### 4.1.4 Adaptation on Another Vowel.



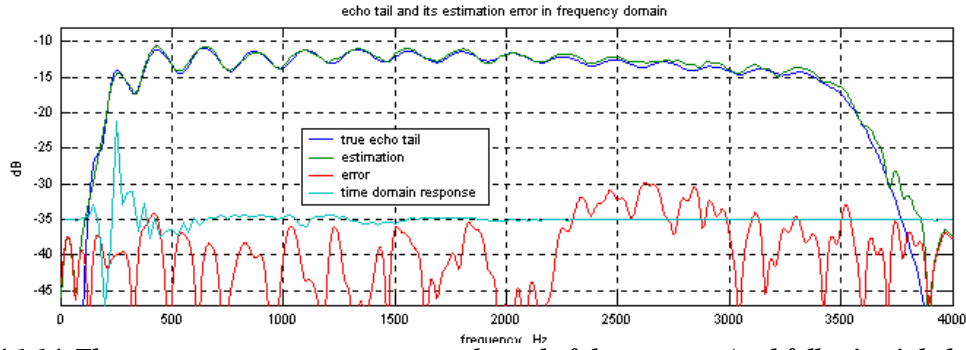
**Figure 4.1.12.** The vowel [ai] is different; the amplitude of the sound is rising and falling. While LEC is adapting to the signal, residual echo (amplified 32 dB) continuously approaches noise-like shape and loses its linear correlation with Snd In signal.



**Figure 4.1.13.** The spectrum of Snd Out signal does not really repeat the shape of Snd In signal spectrum. The frequency maximums (responsible for periodic structure) are suppressed by about 50 dB, but at certain points Snd Out is even higher than Snd In signal. The reason for those discrepancies is simple: the linear part of the echo has been cancelled, and residual echo contains mostly non-linear products of echo path response, which cannot be removed with linear echo path model.

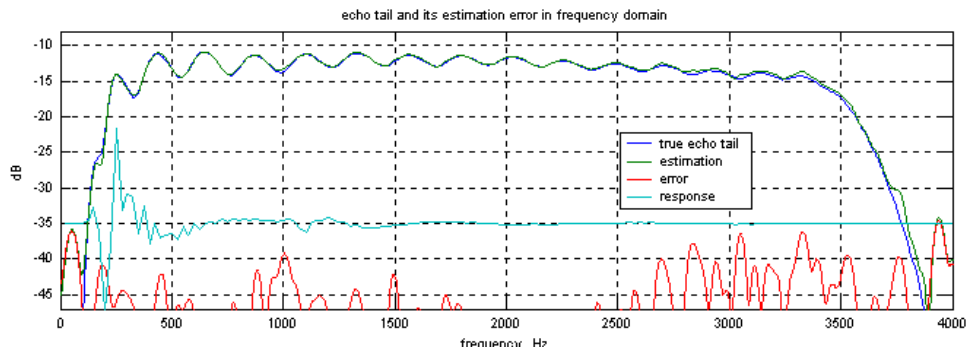
### 4.1.5 Further Behavior

ADF lowers the errors of echo tail estimation by the end of the first sentence.



**Figure 4.1.14.** The convergence error curve upon the end of the sentence (and following inhale) flattens and lowers. LEC is more capable to detect low-level double talk, less and less depending on the current spectral content of voice in Rcv and Snd paths.

The adaptation goes completely to the tracking state in 10...20 seconds with 40% single talk activity in Rcv direction. The echo tail estimation will be fluctuating even in the absence acoustic handset echo, as ADF tries to find out the best linear fit for the current signal.



**Figure 4.1.15.** This is frequency response of a typical echo path error in the tracking state. LEC cancellation ability approaches 30...35 dB in entire frequency range, and the ability to pass double talk transparently is maximal.

The second reflector, delayed by approximately 4.3 ms from the first reflector, is clearly visible now. Its echo is about 15 dB lower. Note the small ripples on the echo path frequency response curve with approximately 230 Hz period. This is in good agreement with the attenuation typically inserted by PSTN and local loops.

## 4.2 Double Talk

These are recordings from another call but in similar conditions. LEC became sufficiently adapted during approximately 10 seconds of single talk, preceding the double talk states to be analyzed here. The near-end speaker is male, and far-end speaker is female. The difference in pitch is almost double, what allows visual distinguishing between residual echo and far-end signals.

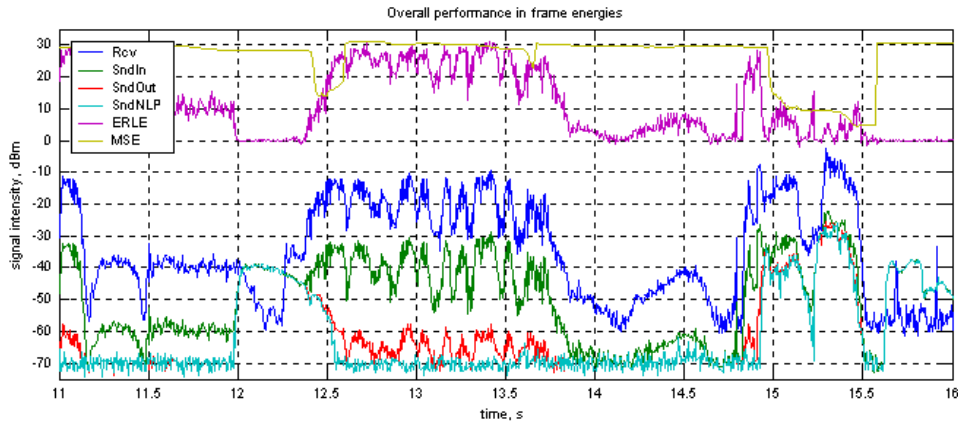


Figure 4.2.1. There are 2 distinct double talk periods at 12 s and 15s from the start of the call.

The performance of LEC can be evaluated by observing the exit from the first double talk and the entrance into the second double talk, when signal amplitude is changing slowly.

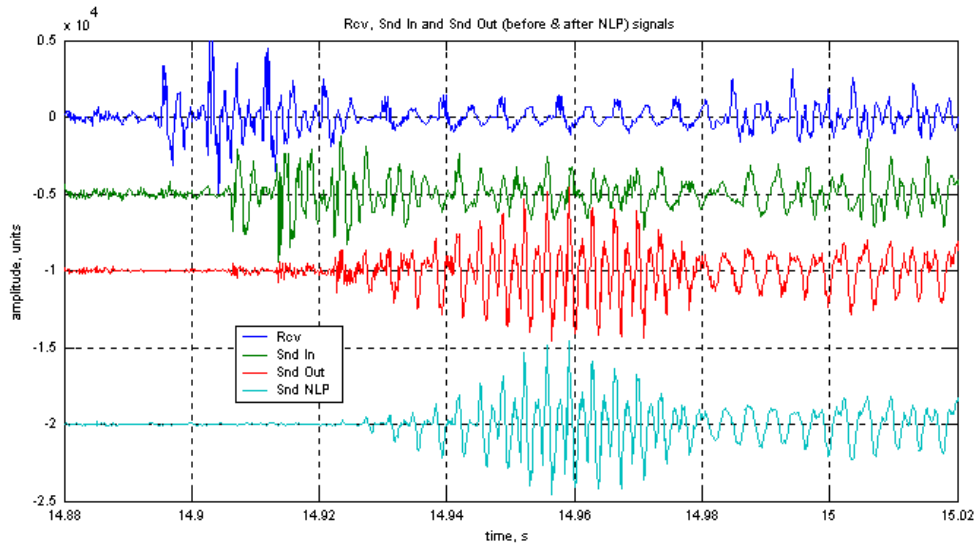
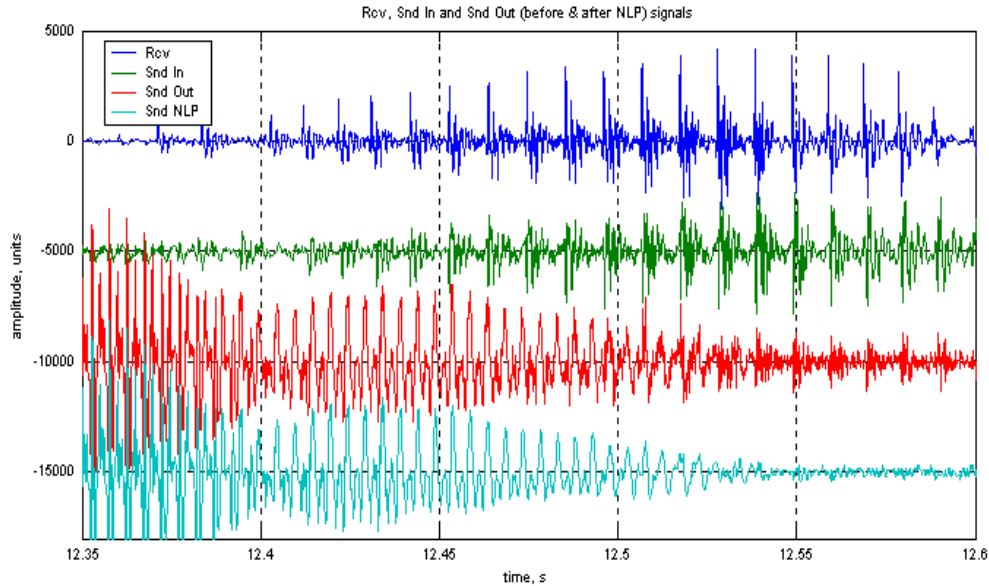


Figure 4.2.2. (SndIn signal is amplified by 20 dB; SndOut signals are amplified by 30 dB). This illustrates slow entrance into double talk state with gradually rising far-end signal. The residual echo signal between 14.91 s and 14.95s contains both echo and far-end signals. NLP removes the remnants of slow-pitched echo by applying post-filtering operations prior to clipping. The post-filtering also improves the output signal between 14.97s and 15.01s during higher-level double talk by removing chaotic remnants of insufficient echo cancellation off the regular curve of voiced far-end signal.

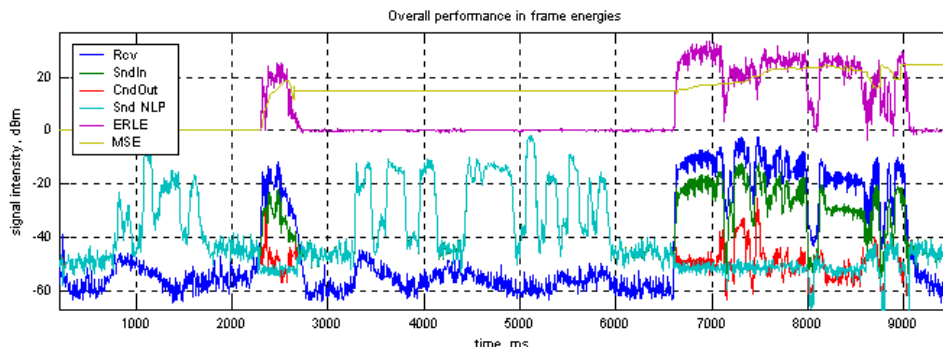


**Figure 4.2.3.** (*SndIn* signal is amplified 20dB; both *SndOut* signals are amplified 40 dB). ADF's output signal contains a mix of echo and far-end speech. NLP lowers the echo component by approximately 10 dB. The output signal of NLP (and LEC as whole) contains little or no visible component with the pitch corresponding to *Rcv* signal.

### 4.3 High Background Noise, Tandem, Low-level Double Talk.

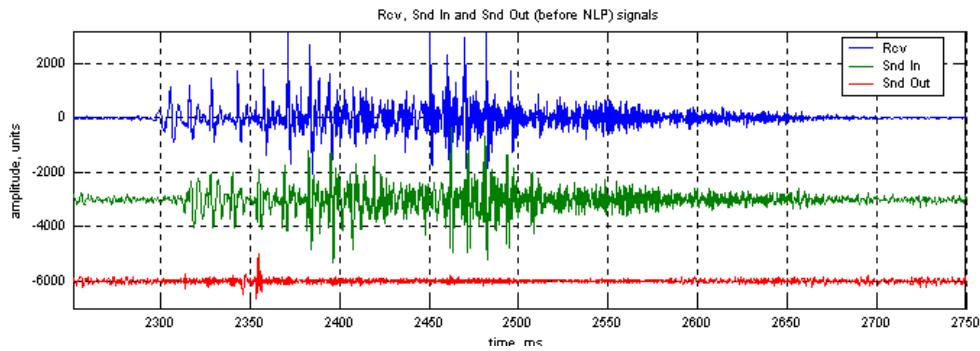
The following pictures illustrate LEC performance in the case of higher background noise. Far end party uses a cellular telephone, which most probably has its own echo controller that displays close-to-half-duplex behavior with sharp noise contrast (LEC residual echo is 5...15 dB during single talk comes to be lower than the level of background noise during pauses). Far end sometimes produces low-level noise bursts.

In the conditions of high-level far-end background noise and tandem with far-end echo controller of questionable quality, LEC is not capable to converge on near-end background noise, breathing sounds and fricatives. Therefore, the overall convergence speed and quality suffer.

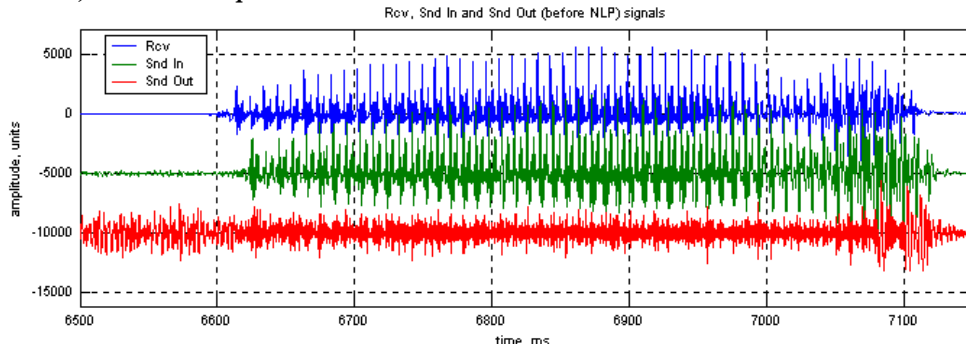


**Figure 4.3.1.** The near-end local loop circuit is known to be imperfect. It distorted sound at 7.1s...7.5s as amplitudes exceed 7500. NLP forcibly removes that excessive echo, as it has been pre-configured to do.

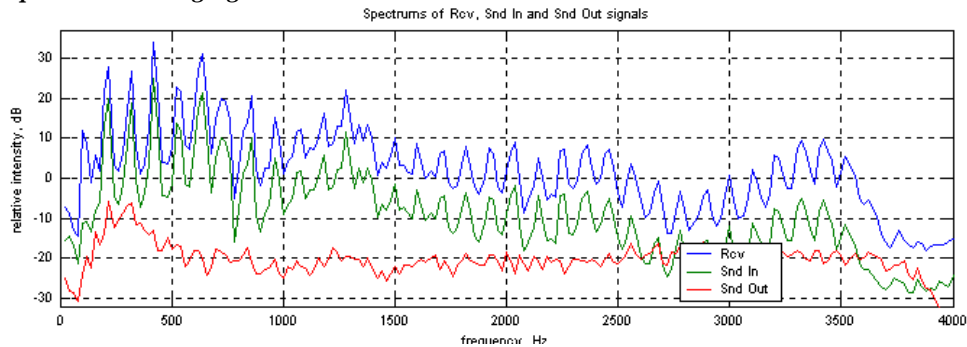
### 4.3.1 Selected Convergence Curves



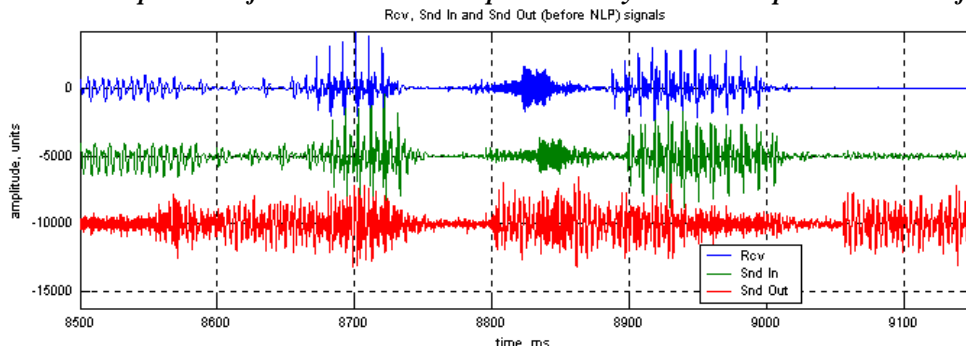
**Figure 4.3.2.** (SndIn and SndOut signals are amplified by 12 dB). LEC displays fast convergence on strong vowels, and removes pitch in 50...100ms.



**Figure 4.3.3.** (SndIn is amplified by 12 dB, SndOut is amplified by 32 dB). The residual echo remains low while pitch is unchanged. The ERLE level is about 30 dB.



**Figure 4.3.4.** The spectrum of residual echo corresponds mainly to the noise produced on the far end.



**Figure 4.3.5.** (SndIn is amplified by 12 dB, SndOut is amplified by 32 dB). The residual echo of the last words shows sharp far-end noise contrast (about 15dB) at the 8.8s and 9.050s time points, most probably corresponding to the end of hangover periods after the end of preceding words. LEC detects low-level double talk at 8.55s and hold it until 9.1s. The SndOut signal (before NLP) during this period is actually pumped up far-end noise.

### 4.3.2 Echo Path Estimation Errors.

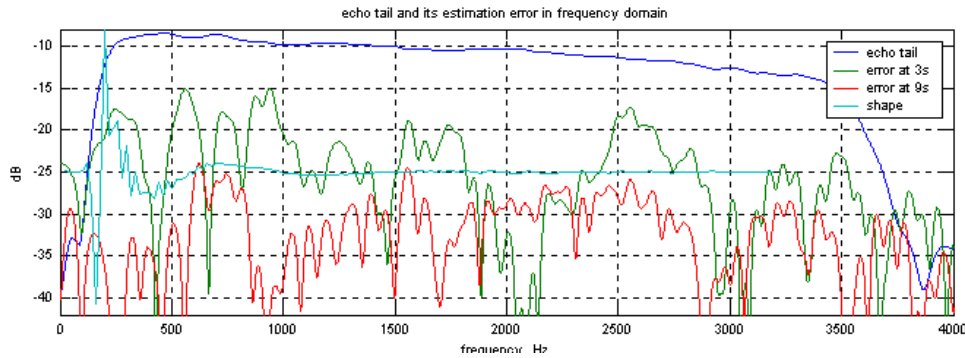


Figure 4.3.6. The echo path estimation errors are decaying but much slower than in the low-noise case. They ADF convergence speed in term of MSE slows proportionally to the background noise energy.

### 4.4 Echo Path Change

A Network EC must perform fast processing on echo path changes because it is not aware of a call start and finish: the start of every call is essentially an echo path change. On the contrary, LEC can be explicitly reset in the beginning (or end) of every call. Far-end echo path change is a rare event. The importance of fast and precise echo path change discovery and reconvergence is much lower for LEC. LEC should keep low the probability of false echo path change detection otherwise double talks may sound odd.

The call, discussed in this case, contained a switching from an attendant to an analog voice mail on the far end, which resulted in the echo path alteration. The speaker was a female using French language.

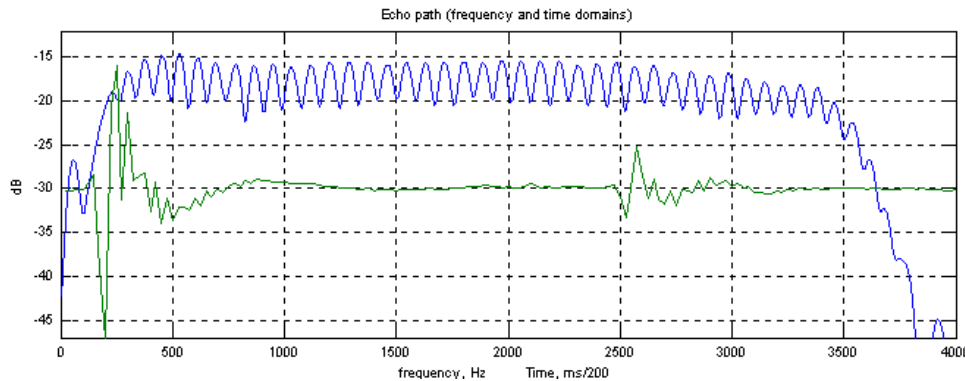


Figure 4.4.1. Time domain echo path before switching contained nothing beyond 10 ms.. The near-end echo has about -17dB level. Most probably, far-end had pure digital connection to far - end CO. As the result of a strong far-end reflector addition (approximately -27 dB level with relative delay of 12 ms), echo path frequency response now exhibits 5 dB ripples all over the pass-band.

LEC was well adapted. Note that if LEC were not fully adapted then the transient period of false double talk would be shorter or it could be eliminated.

The tone, signaling speaker to start recording, appeared at about 50s mark. Then, after a 300ms single tone 387Hz beep (but suddenly for LEC), another reflector appeared in the echo path.

- LEC could not decide between double talk and the echo-path-change during the first sound (50.3 ... 50.7 s, inhale) with relatively low energy. LEC passed the short spike at 50.4s but cancelled the rest. Nevertheless, LEC started to adapt to new conditions and ERLE started to grow immediately after the spike.

- The following words (of formal greeting) had more energy. Initially, LEC assumed double talk condition. LEC switched it off after verifying that the echo is sufficiently correlated with Rcv signal and the echo path is recoverable. False double talk mode lasted for about 0.5 s.
- ERLE climbed above 20 dB by the start on next word, and stabilized on the 30dB level in the next second.
- Higher-than-typical (17dB) ERL added to difficulties in this case because LEC needed sufficiently strong Rcv signal to separate residual echo from background noise.

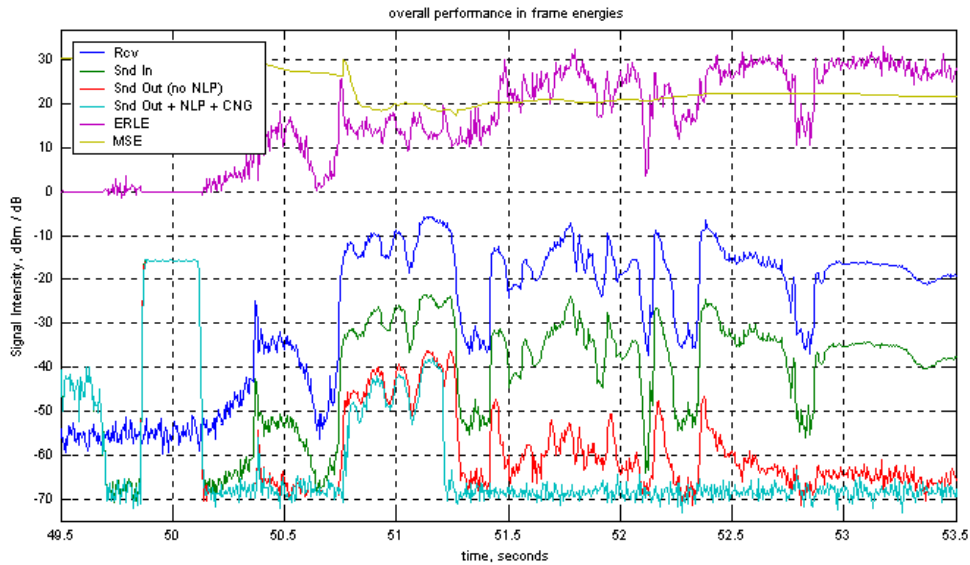


Figure 4.4.2. LEC had accessed MSE as approximately 30 dB just before the switching.

#### 4.4.1 Sound of Inhaling

The time scale on the pictures below is given with reference to 50s mark of Figure 4.4.2.

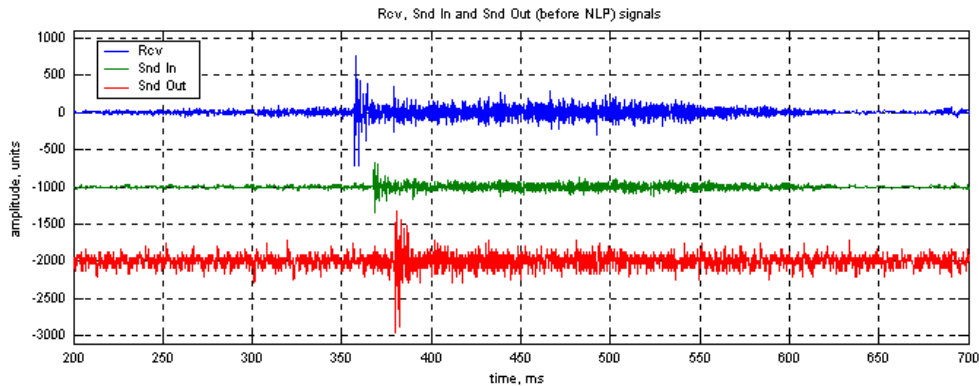
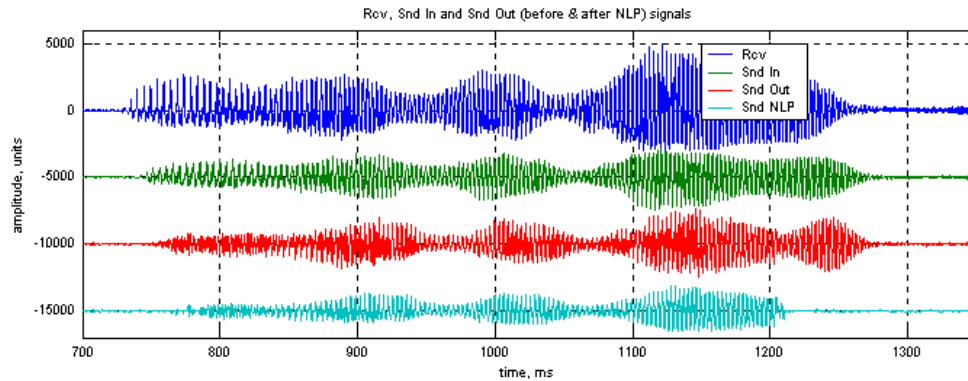
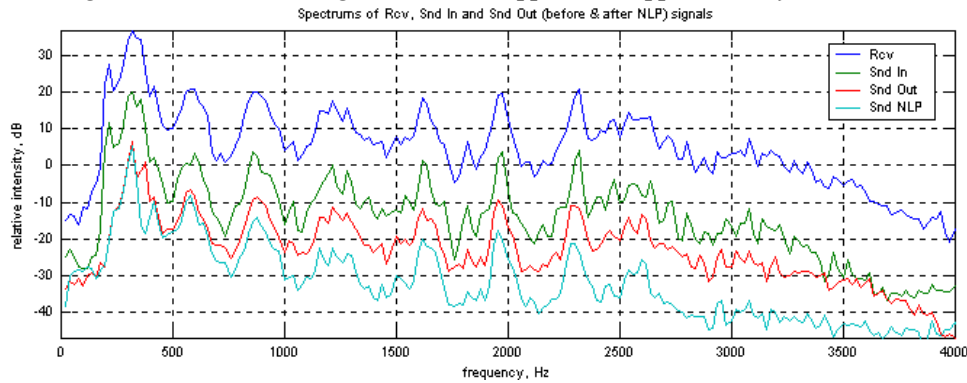


Figure 4.4.3. (SndIn is amplified by 12 dB, SndOut is amplified by 32 dB). The delayed (by approximately 10 ms) under-cancelled echo of inhalation start is clearly visible at approximately 380ms. Thanks to the noise-like spectrum of inhaling sound, LEC could cancel some of the echo, but the quality is still low.

### 4.4.2 Greeting Words

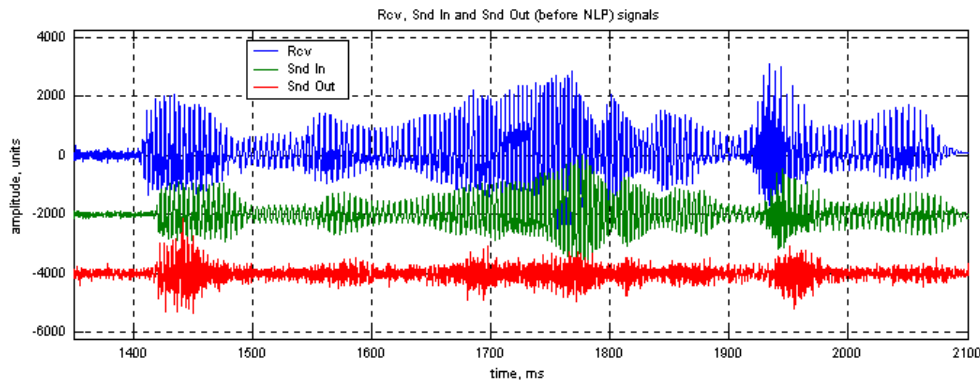


**Figure 4.4.4. Snd In is amplified by 12 dB, Snd Out (both before and after NLP) are amplified by 26 dB. The most of the echo of the following 2 words came back as false echo only slightly modified by NLP. Echo path change has been acknowledged and NLP stepped in at approximately 51.2s.**



**Figure 4.4.5. While converging on the echo path change, LEC kept main ADF filter frozen but NLP post-processed residual echo. The high frequency components became attenuated. This was no longer a distinct hybrid echo, but it sounded as talking to a pipe and it was less annoying. Near-end speaker was surprised of this short echo volume effect from a voice mail system, but continued to talk without interruption.**

### 4.4.3 Following Words



**Figure 4.4.6. SndIn is amplified by 12 dB, and SndOut is amplified by 32 dB. LEC was capable to cancel most of the echo for the next words, effectively removing pitch and producing noise-like output. ERLE is above 20 dB.**

#### 4.4.4 Echo Path Estimation Errors

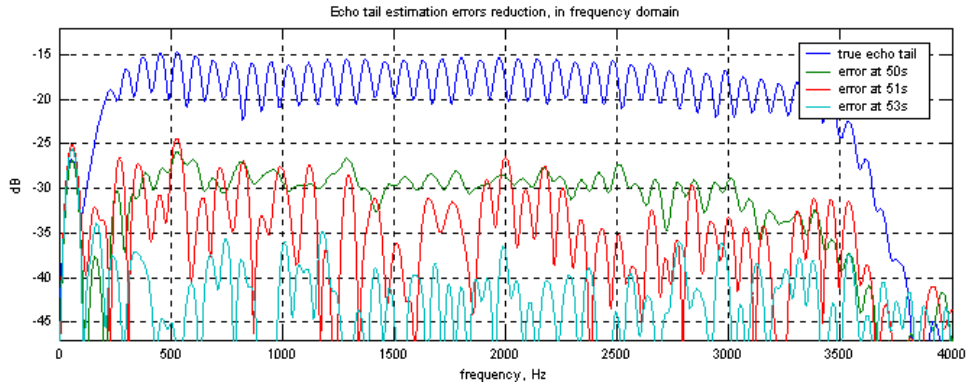


Figure 4.4.7. The echo path estimation errors lower with time. The snapshots of current echo path estimations are taken every second. The monotonically decaying errors are represented in frequency domain.

#### 4.5 Performance on Noise

LEC was not optimized to exhibit particularly good performance on continuous noise.

LEC converges to 20 dB of ERLE within 60...70 ms, what is approximately 3 tail lengths. LEC achieves 30dB of ERLE within 200ms, and improves to approximately 34 dB of ERLE after additional 450ms, when ADF finds the most appropriate adaptation sub-function for the given conditions.

NLP adjusts comfort noise level within 800ms.

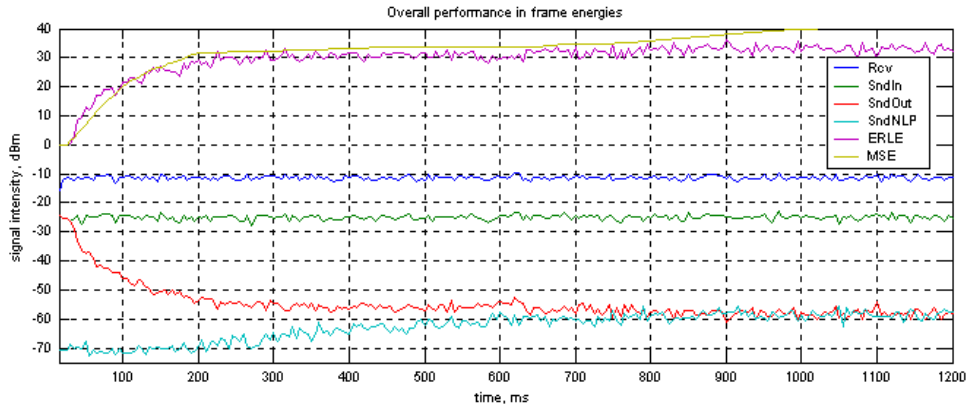
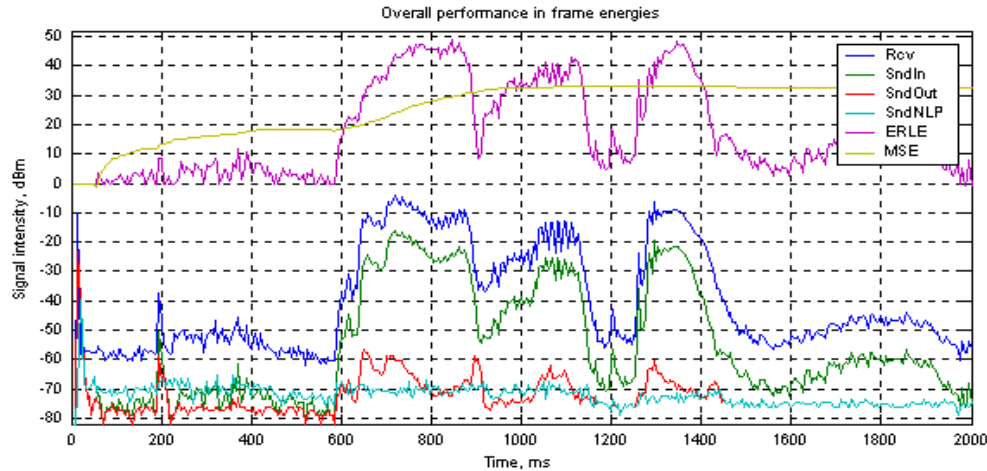


Figure 4.5.1. There are no significant performance advantages of RC-RLS over regular NLMS if Rcv signal is white noise.

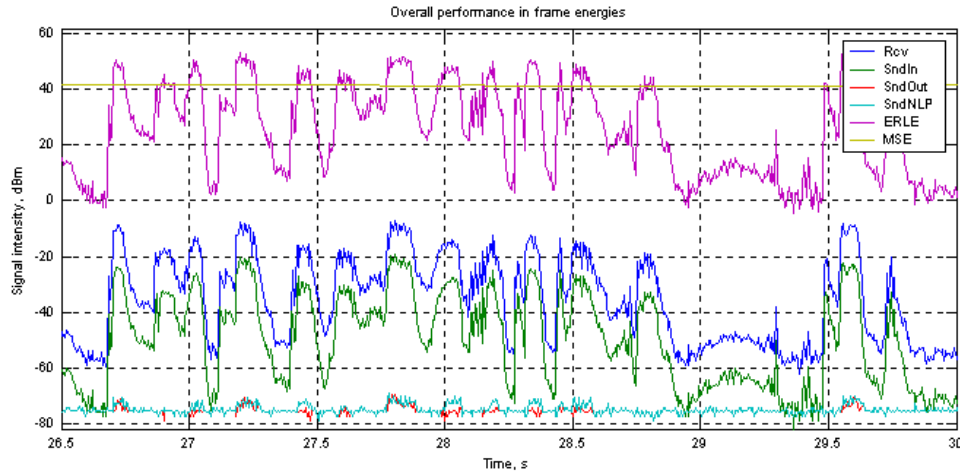
#### 4.6 Performance on Linear Codecs (Estimated).

Entire telephony circuitry and codecs along the echo path are assumed to provide 14 bit linear representation with less than 0.5 bit error. The test signal is the same as of 4.1.

The advantages of linear codecs show off immediately, at the first word.



**Figure 4.6.1. RC-RLS is capable of providing almost immediate echo reduction by 45...50 dB, with corresponding deepening of double talk range.**



**Figure 4.6.2. Eventually LEC becomes able to hide residual echo below -70dB level (the level of occasional plus or minus 1 bit pulses), but that requires about 20...25 seconds of additional convergence time.**

It is important that LEC is preceded by a high-pass / band-pass filter because linear codecs may produce significant DC offset corresponding low-frequency wander. If the codec's Snd filter frequency response does not agree with G.712 requirements, this additional filter must perform adequate corrections. The band-pass filter is design-specific and it is an add-on option for LEC-20 RC-RLS.

## 5 Formal Testing

As mentioned above, neither ITU-T G.165 nor ITU-T G.168 is believed to provide adequate test suites for LEC, but the following data is provided to set correspondence between LEC and G.168-2002 (Draft 6) EC. Tone disabler is not implemented in LEC. Echo path models contained maximum 2 reflectors with less than 15 ms delay (and the first was always at 0 delay). The amplitude of second reflection was lower because the network-inserted attenuation grows as prescribed by ITU-T transmission plan.

No	Test	Result	Notes
2A	Convergence test with NLP enabled	Passed	NLP becomes active immediately.
2B	Convergence test with NLP disabled	Passed	In extended range beyond -30 dB.
2C	Convergence test in the presence of background noise with NLP enabled	Passed	
2C	Convergence test in the presence of background noise with NLP disabled	Passed	
3A	Double talk test with low cancelled-end levels, NLP disabled	Passed	
3B	Double talk test with high cancelled-end levels	Passed	No significant degradations may occur in RC-RLS
3C	Double talk test under simulated conversation	Passed	
4	Leak rate test	Passed	RC-RLS is not leaky by design.
5	Infinite return loss convergence test	Passed	
6	Non-divergence on narrow-band signals	Passed	No degradations because the input signals are recognized as singular and adaptation is disabled.
7	Stability test	Failed	The input sine wave is recognized as singular and adaptation is not enabled. This test is not required for LEC. It is a prerequisite of test 10, which is not applicable.
8	Non-convergence of echo cancellers on specific ITU-T No. 5, 6, and 7 in-band signalling and continuity check tones	Not tested	These signals do not apply to LEC but Network EC only.
9	Comfort noise test	Passed	For all 3 parts
10	Facsimile test during call establishment phase	Not tested	Not applicable
11	Tandem echo canceller test (for further study)	Tested	The results depend on far-end EC.
12	Residual acoustic echo test (for further study)	Tested	The results vary depending on AEC performance by-products. LEC tends to switch to half-duplex mode if AEC does not behave as G.167 prescribes.
13	Performance with ITU-T low bit rate coders in echo path (Optional, under study)	Not tested	
14	Performance with V.Series Low-speed Data Modems	Not tested	Not Applicable
15	PCM offset test (Optional - for further study)	Passed	LEC shall be preceded by a Snd band-pass filter (optional).

## 6 API

### 6.1 Data Structures

#### 6.1.1 Configuration

Configuration is an object to pass control information to LEC instance. The data is copied into LEC database during initialization or a control call. Note that `Int` is defined in TI's supplied `ih` file as 16 bit signed word.

```
typedef struct ILEC_tCfg {
    Int uControl;
    Int sErlMin;
    Int sRcvMax;
    Int sClipThr;
    Int sToneDetectThr;
    Int sToneReleaseThr;
} ILEC_tCfg;
```

The `sErlMin`, `sClipThr` parameters are expressed in 0.1 dB units, so 3.5 Db will be expressed as 35.

The fields of configuration have the following meaning:

Parameter	Description	Recommended Range	Default
<b>Control</b>	Bits masks of commands, defined in 6.3.2.	None	0x0284
<b>ErlMin</b>	Minimum ERL of the echo path. The value is used during initial convergence only.	2...9 dB	3 dB
<b>TxMax</b>	If it is known that analog circuitry is non-linear and this non-linearity appears if the signal exceeds a certain value, fill this field.	None	7770
<b>ClipThr</b>	Threshold used by non-linear processor. The higher threshold is, the more clipping will be audible. The lower threshold is, the more probable is to hear occasional residual echo.	5...10 dB	6.0 dB
<b>ToneDetectThr</b>	The threshold is used by tone auto-detection algorithm. Use higher values if DTMF are likely to be clipped, due to system gain plan.	Q15(0.03)...Q15(0.07)	Q15(0.05)
<b>ToneReleaseThr</b>	The value shall be 2...3 times higher than <code>ToneDetectThr</code> .	Q15(0.07)...	Q15(0.12)

LEC performance depends on the configuration settings. If parameters are set inappropriate, LEC performance will be degraded or LEC may become dysfunctional. Ensure that a user understands the meaning of parameters and consequences of wrong settings before any changes applied.

#### 6.1.2 LEC Statistics

LEC provides following statistics:

```
typedef struct LEC_MIKET_tStts {
    U16 uFlags;
    S16 sTxEn;
    S16 sRxInEn;
    S16 sErrEn;
    S16 sErl;
    S16 sErlE;
} LEC_MIKET_tStts;
```

Parameter	Description	Precision
<b>Flags</b>	Current flags	-
<b>TxEn</b>	Current Tx energy in dB	0.25 dB
<b>RxInEn</b>	Current Rx incoming energy	0.25 dB

Parameter	Description	Precision
RxOutEn	Current outgoing Rx energy	0.25 dB
ErI	Current value of averaged ERL	0.25 dB
Erle	Current valued of averaged ERLE	0.25 dB

Energies are expressed in 0.1 dB units. For detailed information on uFlags meaning see lec\_milet.h

## 6.2 IALG Interface

TI's "Express DSP" compliant interface is fully supported. See relevant TI documentation for further details. Memory allocation details:

No	Element	Referred as	Length (words)	Alignment (words)	Location	Notes
1	Core	pDb	492	2	DARAM	!2, !4
2	Signal History	pHst	480	1	DARAM	!1, !4
3	xDAIS Obj	-	12	2	SARAM/ext	-
4	Scratch-pad	pSc	280	2	DARAM	!1, !2
5	Program	-	4341	1	SARAM/DARAM	! any

Scratch pad is used during LEC\_MIKET\_process(...) invocations. It does not need to be initialized. It may be overwritten by any other application after or before it is used by LEC\_MIKET\_process(...). It may reside in a new place each time a processing function is called: i.e. it may be moved into another location without limitations.

Generally, program shall not share the segments with database or scratch, and each element shall reside in a different segment. Performance will suffer if the database elements (especially signal history or ADF arrays) are placed in internal SARAM. The MIPS-load test program will be provided to ensure that customer-chosen allocation does not lead to severe performance degradation. Consult MIKET DSP Solutions if allocation problems arise.

## 6.3 Vendor Specific API

### 6.3.1 Initialization

```
extern void LEC_MIKET_init_db(
    void *pDb,
    ILEC_tCfg *pCfg,
    Int *pHst);
```

pCfg shall refer to a valid configuration structure.

pDb, pHst shall refer to user-allocated properly-aligned memory blocks.

### 6.3.2 Control

```
extern void LEC_MIKET_control (
    void *pDb,
    Int Cmd,
    ILEC_Status *pStatus );
```

Cmd can constructed by OR-ed flags (other flags will be ignored):

Cmd	Value	Meaning
ILEC_CMD_ON	0x0000	Running by default
ILEC_CMD_OFF	0x0001	Disable LEC from running
ILEC_CMD_RESET	0x0002	Reset LEC
ILEC_CMD_PRESERVE_ADF	0x0004	During reset, preserve adaptive filter from zeroing, to use the same filter for next call
ILEC_CMD_ADAPT_OFF	0x0010	Disable adaptation

ILEC_CMD_POST_OFF	0x0020	Disable post-processing (a part of NLP)
ILEC_CMD_NLP_OFF	0x0040	Disable NLP
ILEC_CMD_LOOPBACK	0x0100	Enforce self-testing loopback
ILEC_CMD_CFG	0x0200	Replace configuration.
ILEC_CMD_TONE_ECHO	0x1000	Remove completely (except for DT condition) echo of the signals, recognized as singular (tone-alike)
ILEC_CMD_DETECT_HIRCV	0x2000	Disable adaptation if Rcv signal exceed Cfg.sRcvMax
ILEC_CMD_CLR_HIRCV_DT	0x4000	Remove non-linear echo produced by excessively strong Rcv signals, even if the current state is double talk.

It is not recommended to apply OFF without RESET due to the obvious consequences of calling `_control(pDb, 0)`; afterwards. Asynchronous calling `_control()` and `_process()` from different tasks is not recommended.

### 6.3.3 Process

```
extern void LEC_MIKET_process(  
    void *pDb,  
    void *pSc,  
    int *pRcv,  
    int *pSnd,  
    int uIsTone);
```

- pDb shall refer to initialized core.
- pSc shall refer to scratch pad.
- pRcv, pSnd shall point to frames (40 samples) of continuous data.
- If it is known that Rcv contains tone, notify LEC explicitly by setting uIsTone flag. Some tones may be clipped and they are very hard to auto-detect.

## 7 Appendices

### Appendix A. Echo Path Properties.

The response of echo path depends mainly on the codecs. Its shape can be easily decomposed from the characteristics of Rx and Tx filters of codecs along the echo path, with taking into account delay between them. The frequency-domain response of analog circuitry is usually wider, although some exceptions apply. The properties of echo tails can be studied in details using cumulative spectral decay (CSD) approach, well developed for clarifying performance details of hi-end acoustic systems.

Echo path is not stationary. There are large and sudden changes when

- phone connects or disconnects,
- a party adds to the conference call,
- people switch phones,
- forwarding happens, as switching from auto-attendant to an extension,
- etc.

LEC needs a period of time to detect the echo path change and to re-converge. Reconvergence will usually happen within first word of sooner, but the exact value depends on the actual voice spectrum, the length of continuous phonemes and sound intensity. The echo will be audible during re-convergence time.

There are also relatively slow and small echo path variations due to the acoustic coupling between a handset's loudspeaker, a cheek, and a microphone. Those variations result in excessive echo, and this excess shall not be misinterpreted as double talk. These acoustic handset echoes are more audible if the handset is not held close to the ear. LEC tracks slow parts of these variations. LEC provides the users with ability to configure the threshold of hiding these non-stationary echoes by NLP. If the threshold is too low, small but fast variations of the echo path may result in some of residual echo going through.

All codecs have limited linearity. A standard  $\mu$ /A-law codec introduces very specific distortions, and they limit the ability of echo canceller to  $\sim 32\dots 35$  dB of echo return loss enhancement (ERLE) even if there is a single  $\mu$ -law codec in the echo path.

Sometimes echo path becomes essentially non-linear. The main reason often is poorly functioning end customer premises equipment.

- Not all acoustic echo controllers are fully complying with ITU-T G.167 and have been exhaustively tested. Many of them are overly simplistic. They may work reasonably against handset on a local call, but do not tandem well with line / network echo canceller, nor with acoustic echo controllers by other vendors (sometimes even the same). They may occasionally produce big bursts of (generally, non-linear long tail) acoustic echo. That may suddenly alter echo path but abruptly disappear later.
- Some feature phones (even from very reputable vendors) produce very bad non-linear echo if used without power supply. Note that the sound going through is still ok, but the non-linear reflections becomes huge and if the peaks exceed the level equivalent to  $-15\dots -10$  dBm.

These cases are beyond the LEC ability to provide decent echo cancellation. LEC will tend to switch into half-duplex mode if those cases happen.

Sometimes, the circuitry in the echo path saturates when the signal exceeds certain thresholds. LEC provides the possibility to configure this threshold and prescribe what should happen if this threshold is exceeded.

Note that, strictly speaking, finite impulse response (FIR) model used by EC for approximation of the echo tail is not physically adequate. There are fewer independent variables than the number of taps in FIR echo path models. Although fractional delay infinite impulse response (IIR) models are closer to the underlying

reality, the adaptive modeling of codec's band-pass filters with poles and zeros almost on unity circle is very dangerous, and fixed-point DSP are much more suited to operations with FIR models.

## Appendix B. NLMS Performance Drawbacks

Any iterative method, used regularly in EC, essentially solves a system of linear equations:

$$h_k = (X_k^T W_k^{-1} X_k)^{-1} X_k^T W_k^{-1} y_k;$$

where  $W_k$  is a correlation matrix of noise (generally, unknown),  $y_k$  is a vector of the signal received from codec, and  $X_k$  is a matrix of signal sent towards the codec (which is, generally, human voice).

Polish scientist Kaczmarz proposed the first iterative scalar step-size algorithm for solving linear systems in 1937<sup>13</sup>. The algorithm was (and still is being) reinvented numerous times. Its most famous reinvention bears the name of Normalized Least Mean Square (NLMS).

$$h_{k+1} = h_k + a_k * x_k * (y_k - x_k^T * h_k) / x_k^T * x_k, \text{ where}$$

$h_k$  is vector of the echo path estimation;  
 $x_k$  is input vector (Rcv);  
 $y_k$  is echo at time k (SndIn);  
 $a_k$  is the step size scalar.

This and many other scalar step-size algorithms have been under intensive studies for many years. The algorithm shall in theory converge ( $trc(E\{var(h_k)\}) \rightarrow 0$  for  $t \rightarrow \infty$ ) for any nonsingular coefficient matrix ( $X_k^T W_k^{-1} X_k$ ), regardless of definiteness, symmetry, or localization of the eigenvalues of the coefficient matrix. In spite of this theoretically stated robustness and the simplicity of the algorithm, the area of its practical applicability is limited and shall not be overstretched.

Let's assume that  $x_k$  is combined of only 2 orthogonal components,  $x1_k$  and  $x2_k$ , (for example, sampled from sine waves,  $f_1 = n * f_2$ ). We can also project  $h_k$  on the same vectors. Then the components will be converging as:

$$h1_{k+1} = h1_k + a_k * x1_k * (y_k - x1_k^T * h1_k - x2_k^T * h2_k) / (x1_k^T * x1_k + x2_k^T * x2_k);$$

$$h2_{k+1} = h2_k + a_k * x2_k * (y_k - x1_k^T * h1_k - x2_k^T * h2_k) / (x1_k^T * x1_k + x2_k^T * x2_k);$$

If the amplitudes of  $x1_k$  and  $x2_k$  are different ( $|x1_k| \gg |x2_k|$ ), then  $h1_k$  convergence will be fast, but  $h2_k$  convergence will be slowed down by factor of  $x2_k^T * x2_k / (x1_k^T * x1_k + x2_k^T * x2_k)$ , what is roughly squared proportion. Lets assume that  $std(x1) = 5 * std(x2)$ . A fixed point NLMS implementation will go through 3 distinct stages:

- EC achieves first 14 dB of echo cancellation (e.g. Echo Return Loss Enhancement, ERLE) on full speed.
- Then the convergence slows down 25 times.
- Then EC stops adapting because the energy-normalized step size for the second mode is so low that the energy-normalized error drops below 1 bit resolution.

As soon as convergence on second mode stops, the convergence on first mode will stop as well because the error term ( $y_k - x1_k^T * h1_k - x2_k^T * h2_k$ ) is combined of both.

Human voice is highly auto-correlated due to the nature of vocal tract excitation. Sound like vowels are periodic, fully auto-correlated, and can be represented via sum of sine waves with frequencies proportional to the pitch (1/repetition-period). The voice signal structure is more complicated than just two sine waves, but the source of the problem is traceable to this very simple case. An EC, based on NLMS, will converge

---

<sup>13</sup> Kaczmarz, S. "Angenäherte Auflösung von Systemen linearer Gleichungen", Bulletin International de l'Academie Polonaise des Sciences, Lett A: 355-357, Cracouie, 1937.

to a 'solution', mostly determined by the current pitch and the position of major formant; or by the frequencies of tones if the input contains DTMF / CP / MF / continuity / etc tones. This NLMS behavior becomes more obvious for acoustic echo cancellation problem.

ERLE may be quite good, but the spectrum of estimation error may be very poor. Low-energy high frequency components may be not cancelled, but they are more perceptually audible. Thus, NLP must be applied to remove the residual echo. If the following input signal has very different spectrum (say, mostly high frequency), NLMS will start to re-converge again, and initially ERLE will be low again. One of two cases will happen:

- The echo of consecutive low-level consonants and fricatives with main energy in high range will be perceived as double talk and their echo will go through NLP;
- Real far-end speech (double talk) will be clipped off by NLP.

No first nor second case nor their combination is desirable. As the result, double talk performance of a NLMS based EC is generally poor.

Summarizing:

- If input is a 'good' white noise like signal, NLMS is an adequate solution. If input is not a 'good' signal, any scalar step algorithm is a variant of a well-paved road to nowhere. Simplified methods are not adequate for high-quality voice echo cancellers.
  - NLMS and other algorithms alike result in the achieving of the flat spectrum of residual error, whatever input signal spectrum is.
  - ERLE is an internal (to EC) measure, but DT range is external, customer-perceivable measure.
  - An EC with deep double-talk range usually has high ERLE value. The opposite statement is not true: an EC with high ERLE may have a mediocre double-talk range.
  - Energies and audibility are not always directly related, due to the properties of human perception system.
  - EC test procedures shall either avoid using noise –like signals, or avoid drawing far-stretching conclusions from such experiments.
-

DISCLAIMER OF WARRANTIES; LIMITATION OF LIABILITY:

MIKET DSP SOLUTIONS HAS MADE EVERY REASONABLE EFFORT TO ENSURE THE ACCURACY OF THE DATA, HOWEVER, MIKET DSP SOLUTIONS SPECIFICALLY DISCLAIMS ANY WARRANTY, EXPRESSED OR IMPLIED, RELATING TO THE PRECISION OF THE DATA PROVIDED, THEIR COMPLETENESS OR QUALITY, INCLUDING ANY IMPLIED WARRANTY OF FITNESS FOR ANY PARTICULAR PURPOSE. CUSTOMERS ARE RESPONSIBLE FOR THEIR APPLICATIONS USING MIKET DSP SOLUTIONS COMPONENTS, AND AGREE THAT MIKET DSP SOLUTIONS SHALL HAVE NO LIABILITY WHATSOEVER FROM ANY CLAIM OF LOSS OR DAMAGE OF ANY ALLEGED ERROR OR DEFECT. IN NO EVENT SHALL MIKET DSP SOLUTIONS BE LIABLE FOR INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. IN NO EVENT SHALL MIKET DSP SOLUTIONS' LIABILITY, INCLUDING FOR DIRECT DAMAGES, EXCEED THE AMOUNTS PAID IN CONNECTION WITH LICENSING OF MIKET DSP SOLUTIONS' COMPONENTS.

---

ALL TRADEMARKS ARE PROPERTY OF THEIR RESPECTIVE OWNERS.