ACOUSTIC ECHO CANCELLER
(GAEC)
Data Sheet

The data contained in the document is preliminary and is subject to change without notice.
TABLE OF CONTENTS

SUMMARY ............................................................................................................................... 4
ABBREVIATIONS .................................................................................................................. 5

1 INTRODUCTION .................................................................................................................. 6
1.1 THE SOURCE OF ACOUSTIC ECHO ................................................................................... 6
1.2 HUMAN EAR SENSITIVITY TO DELAYED ECHO ................................................................ 6
1.3 BASIC AEC .......................................................................................................................... 7
1.4 ADDING ECHO CANCELLATION TO A BASIC AEC ........................................................... 8
1.5 NEED FOR QUALITY SPEAKERPHONES ........................................................................... 10
1.6 IMPORTANCE OF GOOD ACOUSTICS IN A SPEAKERPHONE ......................................... 11

2 THEORY OF OPERATIONS .................................................................................................. 12
2.1 MAIN PRINCIPLES ............................................................................................................. 12
2.1.1 Adaptive filtering ........................................................................................................... 12
2.1.2 Post-processing ............................................................................................................. 12
2.1.3 Noise reduction ............................................................................................................. 14
2.2 PROCESSING BLOCKS ....................................................................................................... 15
2.2.1 Receive path processing ............................................................................................... 15
2.2.2 Send Path processing ................................................................................................... 16
2.2.3 Adaptation .................................................................................................................... 16
2.2.4 GAEC control ............................................................................................................... 16

3 API ......................................................................................................................................... 17
3.1 DATA STRUCTURES ............................................................................................................ 17
3.1.1 Configuration ................................................................................................................ 17
3.1.2 GAEC statistics ............................................................................................................. 18
3.2 IALG INTERFACE .............................................................................................................. 18
3.3 VENDOR SPECIFIC API .................................................................................................... 19
3.3.1 Initialization .................................................................................................................. 19
3.3.2 Control .......................................................................................................................... 19
3.3.3 Process .......................................................................................................................... 20

4 INTEGRATING GAEC INTO A SPEAKERPHONE .................................................................. 21
4.1 INTERFACING AND DISTORTION CORRECTION FILTERS .............................................. 21
4.2 MIPS AND MEMORY ......................................................................................................... 22
4.2.1 Memory consumption ................................................................................................... 22
4.2.2 MIPS on C54x ................................................................................................................ 23
4.2.3 MIPS on C55x ................................................................................................................ 23

5 SPECIFICATIONS .................................................................................................................. 24
5.1 TRANSMISSION SPECIFICATIONS ................................................................................. 24
5.2 ACOUSTIC ECHO CONTROL SPECIFICATIONS ................................................................. 24
5.2.1 Measurement signal ....................................................................................................... 24
5.2.2 Acoustic echo path ......................................................................................................... 24
5.2.3 Weighted terminal coupling loss – single talk (TCLwst) ................................................... 26
5.2.4 Weighted terminal coupling loss – double talk (TCLwdt) ............................................... 26
5.2.5 Received and sent speech attenuation during double talk (Ardt/Asdt) ............................. 27
5.2.6 Received and sent speech distortion during double talk (Drdt/Dsdt) ............................. 27
5.2.7 Maximum frequency shift ............................................................................................. 27
5.2.8 Break-in time – single talk (Tonst) ................................................................................. 27
5.2.9 Break-in time – double talk (Tondt) ................................................................................ 28
5.2.10 Initial convergence time (Tic) ...................................................................................... 28
5.2.11 Recovery time after double talk (Trdt) .......................................................................... 29
5.2.12 Terminal coupling loss during echo path variations (TCLwpv) ........................................ 29
5.2.13 Recovery time after echo path variations (Trpv) ................................................................ 30

5.3 SPECIFICATION FOR INTERWORKING WITH THE NETWORK ........................................... 30
5.3.1 Interworking with speech codecs ...................................................................................... 30
5.3.2 Interworking with network echo cancellers ....................................................................... 30
5.3.3 DCME and PCME interworking ......................................................................................... 30
5.3.4 Interworking between a wide-band terminal and other types of terminals through the network. 30

5.4 ADDITIONAL SPECIFICATIONS .......................................................................................... 31
5.4.1 Convergence for various RCV IN levels .......................................................................... 31
5.4.2 Double talk sensitivity ....................................................................................................... 31

6 CHARACTERIZATION OF PERFORMANCE ON VOICE SIGNALS ........................................ 34
6.1 SINGLE TALK CONVERGENCE ............................................................................................ 34
6.2 DOUBLE TALK DETECTION AND PROCESSING .................................................................. 35
6.2.1 Double talk in early stage of convergence ...................................................................... 36
6.2.2 Double talk on the 2nd repetition .................................................................................... 37
6.2.3 Further improvements ....................................................................................................... 39

6.3 SPEAKER CHANGE .............................................................................................................. 41

7 CHARACTERIZATION OF PERFORMANCE IN A SPEAKERPHONE ..................................... 43
7.1 SETUP #1 .............................................................................................................................. 43
7.1.1 Convergence on noise ....................................................................................................... 43
7.1.2 Convergence on voice ....................................................................................................... 43
7.1.3 Subjective testing .............................................................................................................. 45
7.2 SETUP #2 .............................................................................................................................. 45
7.2.1 Convergence on noise ....................................................................................................... 46
7.2.2 Convergence on voice ....................................................................................................... 47
7.2.3 Subjective testing .............................................................................................................. 47

7.3 CONCLUSIONS .................................................................................................................... 48
Summary

Due to acoustic reflections, the signal from a loudspeaker’s microphone contains a mix of the acoustic echo of far-end signal and the signal of near-end party. The purpose of acoustic echo control in a speakerphone is to separate these signals and eliminate acoustic echoes, i.e. to make the echo perceptually inaudible, while preserving the signal of near-end party as much as possible\(^1\).

- The proposed G.167 compliant acoustic echo controller GAEC is capable of providing nearly full duplex quality in average-sized office rooms when used in inexpensive speakerphones with single omni-directional microphone.

- GAEC is built around a single DSP platform (TI’s `C5000 family DSPs). GAEC is primarily designed for VoIP/VoLAN/PBX phones, which already employ a DSP to perform signal processing functions. GAEC can be configured not to occupy all MIPS in the DSP and leave enough resources to perform phone control functions and other activities.

- GAEC incorporates results of recent technological advances in transform domain multirate adaptive processing, psychoacoustics, adaptive post-filtering, noise reduction, sensitive and robust echo path change and double talk detectors into an integrated DSP algorithm.
  - GAEC achieves up to 50 dB of terminal coupling loss (TCLwst) by adaptive means, without inserting any attenuation, with initial speed exceeding 100 dB/s.
  - GAEC adapts in wide dynamic range, even on signals with energy as low as –60 dBm.
  - GAEC provides double talk range (both detection and processing) close to TCLwst.
  - GAEC performs spectrum weighting to support the least distorted, perceptually transparent sound in difficult double talk conditions.
  - GAEC minimizes signal clipping during activity switching.
  - GAEC supports nearly transparent mode during inactivity periods, preserving natural-sounding background noise.
  - GAEC reduces background noise levels by up to 15 dB.
  - GAEC works with echo tails up to 400 ms.
  - GAEC is fully capable of smooth interoperating with G.168 network echo cancellers.

- GAEC accounts for loudspeaker non-linearity and is capable of performing in full dynamic range even with inexpensive loudspeakers, suffering from significant distortions.

- GAEC can be used for 2-wire (residential) phones. A line echo canceller (which can be supplied by MIKET DSP Solutions) shall be added to the design to combat hybrid echo (from microphone to loudspeaker) and to compensate for GAEC’s 6.25 ms of delay in each direction.

- Although GAEC is not specifically tailored for teleconferencing with many people in the same big room, a speakerphone equipped with GAEC can be used for teleconferencing if the people are not dispersed too far from each other and the room’s reverberation time is below 700ms. The GAEC quality will smoothly degrade with the increase in the room size.

- GAEC can be extended to operate with more than one microphone and perform active beam forming in real time along with acoustic echo control operations if ‘C55x platform is used, what significantly increases speakerphone’s quality by room “de-reverberation” and possible exclusion of some of the room resonance modes.

\(^1\) ITU-T Recommendation G.167. “Acoustic Echo Controllers” (07/93).
**Abbreviations**

AEC  Acoustic Echo Controller (or Canceller)
DSP  Digital Signal Processing, as the theoretical base of the AEC operations
DSP  Digital Signal Processor, a specialized processor for performing math intensive operations.
DT  Double Talk condition, when both far end and near end parties are talking at the same time.
DTD  Double Talk Detector
ERL  Echo Return Loss, a characteristic showing how the echo energy is related to RCV OUT energy
ERLE  Echo Return Loss Enhancement, a characteristic showing how well adaptive filters perform echo cancellation
FE  Far End party, not present in the same room as the speakerphone
NE  Near End party, speaking into speakerphone’s microphone
NLP  Non Linear Processor, a common name for the processing of residual echo after echo cancellation
RCV  Receive direction, towards the loudspeaker’s loudspeaker
SND  Send direction, from the loudspeaker’s microphone
TCL  Terminal Coupling Loss
1 Introduction

1.1 The source of acoustic echo
Acoustic echo is formed when the sound emitted by a speakerphone’s loudspeaker gets reflected from the walls, ceilings, floor, furniture, people, etc. back to the speakerphone’s microphone. Sound pressure level decreases with each reflection. Some surfaces, as heavy carpet, soft furniture, open half-full bookshelves with varying format books in random order, people, animals and especially acoustic foam and panels, reflect very little but absorb, dissipate or otherwise significantly attenuate acoustic echo. Surfaces as glass, brick, gypsum board walls, etc reflect about 95% of the sound back. The reflections, being repeated multiple times, create reverberation effect.

Typically, the reverberation level decreases exponentially with time, so the rooms are often characterized by $T_{60}$, which specifies the time when reverberation level drops by 60 dB ($T_{30} = T_{60}/2$). For a typical office, $T_{60}$ lies between 300 and 600 ms.

The binaural human hearing system copes remarkably well with reverberation effect when both people in conversations are in the same, even highly reverberant, room. This is not the case if the same people are in different rooms and they use a speakerphone for conversation. High-level acoustic echo becomes very annoying and disturbing, and thus it shall be removed to enable handset-like conditions. Moreover, if both people use speakerphones, the acoustic feedback may (and often does) lead to ringing and howling, thus disabling the conversation entirely.

With nominal gains in the loudspeaker’s and microphone’s path, the acoustic echo is usually about -8…-3 dB in typical rectangular plain-wall office rooms without much of soft furniture. The resonating cabinet of the speakerphone itself usually drives the total echo up to +5…+10 dB. In some conditions, acoustic echo may be 20 dB higher than the original received signal.

1.2 Human ear sensitivity to delayed echo

![Diagram](image)

Figure 1.1. The absolute threshold of perceptibility of a delayed single echo of on human speech decreases as 0.5 dB per ms (approximately).

---

2 Room acoustic is a complicated issue, but it has been extensively studied in depth from both theoretical and practical perspectives. See specialized literature, as H. Kuttruff, “Room acoustic”, Applied Science Publishers Ltd, London, 1973, for further and deeper description.

3 Researches note that the sound processing by brain plays the major role in the human hearing system. See: Peter H.Lindsay, Donald A. Norman. “Human information processing”, University of California, San Diego, Academic Press, NY and London, 1972.
Usually, the echo is most audible on transitions between phonemes and especially to silence, and least audible on long vowels. The perceptibility and annoyance of the echo is influenced by many subjective and objective factors as language spoken, emotional context, importance of particular conversation, hearing skills, etc.

The major factor in the delayed echo perceptibility is the delay time between direct signal (as far-end handset’s sidetone signal) and the echo and their relative strengths. Figure 1.1 illustrates the rough ‘rule of thumb’ for the assessment whether the echo is audible, assuming 0 dBm as the level of the sidetone in far-end handset.

1.3 Basic AEC

The simplest approach to the acoustic echo control is to switch some attenuation between send (SND) and receive (RCV) paths.

![Diagram of Basic AEC block diagram.](image)

Note that mild attenuation (less than 10 dB flat over frequency range) would allow both parties to communicate more or less freely, but any attenuation higher than 25 dB effectively enforces half-duplex operations and automatically gives priority to far-end.

The attenuation, inserted by such basic AEC to make the acoustic echo inaudible, shall account for:

- The maximal strength of the acoustic feedback;
- RCV path volume setting (amplification or attenuation);
- Maximal expected delay on the connections between far-end and the loudspeaker, and ability to interwork with Network Echo Cancellers on long distance calls.

To ensure that the cumulative attenuation between RCV IN and SND OUT, also referred as Terminal Coupling Loss (TCL), is high enough (40...50 dB), AEC shall switch about 45...60 dB of attenuation between SND and RCV.

The acoustic echo level is high, thus it is nearly impossible for a basic AEC to distinguish between the echo and true near-end signal. So AEC significantly attenuates the SND path signal whenever RCV is active. When SND activity ends, SND attenuation is gradually switched off to RCV, according to the reverberation time $T_{60}$ for the target room sizes.

---


Copyright 2001-2003 MIKET DSP SOLUTIONS, All rights reserved.
The near-end party shall always patiently wait till the far-end party ends speaking – or speak very loudly (nearly shout) into the microphone to enforce AEC to switch SND path attenuation off. That is clearly asymmetric and puts the party on the speakerphone to disadvantage in critical conversations, and sometimes leads to conflicts, which could be avoided if the conversations over speakerphone were full duplex.

Some of the other disadvantages of the basic AEC are:

- Switching to the far-end party is often abrupt, and the starting unvoiced phoneme may be easily clipped, so “four” may sound as “or”, etc.

- Due to the high attenuation inserted, the less active channel becomes totally silent. In the countries, where the end of the call is not indicated by call progress tones, the near end may continue talking in void, while far end party has disconnected a long ago.

- A typical handset exhibits frequency response ±1…3 dB off nominal in the pass band, in both directions. If no special care is taken, an uncorrected speakerphone may exhibit ±15 dB deviations off the nominal in its pass band, in either direction, thus significantly distorting voice sounds.

- The average distance between mouth and microphone is assumed to be about 2’ for the speakerphone case instead of couple of inches for handset. So the microphone circuitry shall provide 15…30 dB more amplification, what amplifies undesirable background noise and decreases intelligibility of weak consonants in the near end speech.

- It is nearly impossible to provide microphone directivity in a simple design with a single inexpensive, usually omni-directional, microphone. So the far-end party hears the near-end voice being ‘enhanced’ by the room reverberation. That might be good for an opera singer audition over a phone, but in other situations it is quite annoying.

It is no surprise that most people do not like using speakerphones.

**1.4 Adding echo cancellation to a basic AEC**

The most obvious way to improve the basic AEC is to add to it an adaptive filter, the same way it is done in Network Echo Cancellers (Figure 3).

![Figure 1.3. Basic AEC can be enhanced by adaptive filtering.](image-url)
Adaptive filter (ADF) simulates the echo path response in linear approximation. Adaptive filter is convolved with the RCV signal history (HST) to obtain the echo estimate, which is then subtracted from SND IN signal. This process is called echo cancellation. The resulting error signal is used to tune adaptive filter with algorithms ranging from NLMS to RLS.

If adaptive filter provides X dB of echo cancellation (an adaptive part of TCL), then the amount of attenuation switched between SND and RCV (a switched part of TCL) can be lower by the same X amount. In theory, if adaptive filter can provide echo cancellation equal to the target TCL figure, the switching becomes unnecessary and an AEC starts to operate in true full-duplex mode.

This approach, being theoretically quite sensible, does not lead to significant improvements in practice, due to the variety of reasons:

- The adaptive filter needs to be very long to provide at least 30 dB of echo cancellation. Rough estimate is \(0.5 \times T_{60}\), what corresponds to 200…400 ms and equals to 1600…3200 taps in the filter. That stipulates very high MIPS load on DSP (tens or hundreds of MIPS). The longer echo tail is, the more challenges it represents for 16-bit fixed-point implementations, and the closer it approaches the point when adaptation becomes theoretically impossible in fixed-point arithmetic.

- The echo path linearity assumption shall not be overstated. Although the acoustic reflections are quite linear for reasonable sound pressure levels, a typical speakerphone’s loudspeaker is very far from being a linear transducer\(^5\). That gives rise to many problems and puts a hard limit on the depth of echo cancellation.

- The speakerphone’s acoustic echo path is not stationary and varies with slightest moves of the people in the same room\(^6\). The longer AEC’s echo tail is, the more pronounced is this effect. A Calman filter, with its elaborated step size control, is indeed required for the AEC, as well as echo path change and double talk detectors.

- The length of the required AEC’s echo tail (200…400ms) greatly exceeds the duration of non-voiced speech segments (10…30ms). It shall be noted that the speed and quality of full-band convergence significantly suffers in this case, and in some cases the convergence becomes questionable due to the singularity of the voiced speech segments as the excitation signal for adaptive filters.

As the result, simple adding of adaptive filtering to basic AEC may only result in a modest decrease of the amount of attenuation switched between RCV and SND, as 10…15 dB. That does not solve the problem because, from the user’s perspective, there is no perceptible difference between switching 45 dB and 35 dB – both values are too high.

Annex A of ITU-T Recommendation P.340\(^7\) contains results of the double talk performance tests in dependence on the attenuation inserted in the SND path. Overall quality MOS decreases to 3.0 if the attenuation is only 6dB, to MOS 2.0 if the attenuation is 12 dB, and MOS comes to 1.0 if the attenuation exceeds 24 dB.

Some users find that a simple, but well-tuned acoustic echo controller outperforms adaptive echo cancellers.

---


1.5 Need for quality speakerphones

If regular speakerphones were of good quality, people would use them frequently just because of hands-free convenience. Obviously, this is not the case. But, there are many situations when the need for a quality speakerphone is quite apparent, as for

- teleconferences, with many people at the same office required to communicate to the same far-end party;
- situations, when a near-end party requires 'handsfree' mode for performing various operations in parallel with talking, but wired headset is restrictive and /or inappropriate;
- people like sales agents or managers, who talk on the phone each day for long time, and soon become tired of using handsets;
- video-conferencing, which is awkward with either handsets or headsets.

The problem with teleconferencing was usually solved with expensive ($700-$1000) conferencing phones, whose price used to be well justified because:

- A network of DSPs was required to perform adaptive processing for an average conferencing room.
- Conferencing phones may use multiple microphones to ensure more uniform pick-up for all participating people.
- Conferencing phones use high quality loudspeakers, able to provide sufficient sound pressure level with minimal distortions, and have carefully crafted housing to soften internal resonance modes.

The conferencing phones perform significantly better than a basic speakerphone, but their merits are offset by their high price, which practically disables their usage for any other purpose than solely teleconferencing from the company’s boardroom.

There were no viable alternative solutions till recent times, when DSPs became much more powerful and affordable, and are now capable of performing quality AEC operations in a single entry-level DSP for total cost of less than $10, also leaving enough space (in terms of MIPS an memory) for performing phone line and button control functions.

Moreover, many improvements over basic adaptive AEC have been proposed in the course of last 20 years, including frequency$^8$ (or transform / sub-band / Gabor$^9$) domain adaptive filters$^{10}$, various post-filtering algorithms$^{11,12}$, ways for exploiting human ear properties$^{13}$, etc. Yet, very few, if any, of those theoretical advances have reached the end user. The source of the problem can be attributed to the big, and seemingly increasing with time, gap between academic research and practical implementations. A practical implementation must incorporate very many theoretical advances from different sources altogether to achieve customer-visible improvement over the basic a AEC to provide high quality, nearly full-duplex connection. There is no one “magic” algorithm, which is solely capable of closing this gap.

The proposed GAEC$^{14}$ uses recent advances in both DSP theory and DSP silicon, and is capable of providing an affordable quality alternative to expensive conferencing phones at much lower cost, especially when used in a speakerphone with reasonably good loudspeaker and acoustically adequate housing.

---

14 The kernel of GAEC’s proprietary complex domain multirate adaptive filter is based on the recent advances in bi-orthogonal perfect reconstruction filter banks. Surprisingly enough, they (in a big part) represent a re-discovery of a principle proposed by D. Gabor in 1946.
1.6 Importance of good acoustics in a speakerphone

It shall be noted that a good AEC is only a half of the solution. If a speakerphone’s acoustical quality is very poor, it is unlikely that any AEC solution can help to produce a full-duplex sound:

- If no special care was taken of acoustical dumping then standing waves (internal to speakerphone resonance modes) can develop in the 1000…3000 Hz range \(0.5 \times \lambda_{1000Hz} = 0.16m\), what is close to the size of a typical speakerphone.
  - These stand-waves ‘pull’ the energy from ‘outside’ to ‘inside’, and the speakerphone’s microphone often becomes excessively coupled to the loudspeaker.
  - As the result, the transmission curve has a loss of 5…10 dB in the range, and the acoustic coupling has a gain of roughly equal strength.
  - To compensate for the transmission loss, the speakerphone shall employ an equalizing RCV filter, which doubles the coupling effect for the far-end.
  - Thus an AEC must switch extra 10…20 dB of attenuation to preserve TCL, what shifts the speakerphone towards half-duplex mode.

- Inter-modular distortions of the loudspeaker put a hard limit on the maximum achievable TCL. The higher are the distortions, the lower is the adaptive part of TCL, and the more attenuation AEC shall switch between RCV and SND.

- Most users would increase RCV volume if a loudspeaker has significant inter-modular distortions (just to hear the far-end clearer). Then an AEC solution must increase switched attenuation to preserve TCL figure for the far-end, what contributes to the shift towards half-duplex mode.

- Some cheap loudspeakers are so non-linear, that the echo path as perceived by an AEC does not fit into linear approximation anymore. Sufficient TCL figure may be achieved by attenuation only, therefore only half-duplex sound can be expected from these terminals.
2 Theory of operations

2.1 Main principles

Three of the main principles of GAEC operations are discussed in this chapter on a high abstraction level.

Further details are beyond the scope of this document.

2.1.1 Adaptive filtering

GAEC uses a complex domain generalized sub-band adaptive approach, coupled with sub-optimal adaptation speed control, to achieve up to 40 dB of Echo Return Loss Enhancement (ERLE) with maximal speed of about 100 dB/s.

The main advantages of the sub-band approach are:

- Each sub-band’s adaptation process can be independently controlled, and the building of appropriate sub-optimal step-size control algorithm is simplified for the narrow sub-bands.
- Human voice pitch tends to vary (depending on the emotional context) during the call. After sub-band decimation, the eigen-value spectrum of the decomposed signal becomes spread and consequently the task of adaptation, especially in higher sub-bands, becomes easier.
- The length of adaptive sub-filter in each sub-band is shorter.
- The variability of the ERL frequency response is much lower within each individual sub-band.
- MIPS can be significantly decreased.

The 40dB limit on ERLE was set as a compromise between filter bank delay, MIPS and memory resources of modern affordable DSPs, the expected loudspeaker non-linearity, and echo path variability during a typical conversation.

GAEC can adapt on background noise and achieve up to 10 dB of potential echo cancellation in 200 to 500 ms, even before the first word is pronounced.

GAEC uses normalized (“block-floating-point”) representation of main signals and extends the dynamic range of adaptation down to the –55...-65 dBm range. The 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.

GAEC’s adaptive filter was designed for human voice excitation, it takes into account that high-energy voiced vowels tend to have singular or very poor eigen-spectrum distribution, what makes convergence difficult, but the signals of low-level fricatives, consonants, breathing, and background noise have smoother spectrum than the spectrum of strong periodic vowels.

The further details of the principles of sub-band adaptive filtering can be found in many publicly published articles (as proceedings of ICASSP conferences). The particulars of GAEC adaptive filter are beyond the scope of this document.

2.1.2 Post-processing

Any adaptive filter by itself, however advanced, is debatably capable of providing sufficient de-coupling (target TCL) between RCV In and SND OUT even in nearly ideal conditions. There always is excessive residual echo, either linearly related to RCV signal, due to the echo path non-stationary nature, or non-linearly, due to the loudspeaker non-linearity.

Instead of switching attenuation between SND and RCV to achieve the required degree of decoupling between SND and RCV, GAEC uses a proprietary post-processing technique, which is based on the properties of human voice production and perception.
Humans recognize the phonemes by analyzing so-called formants, or peaks in short-term spectrum, and the properties of signal between those formats are less important for correct phoneme recognition (what is exploited in various low bit-rate LPC and CELP codecs as LD-CELP, VSELP, etc). If the RCV IN – SND OUT de-coupling problem is formulated as a short-term spectrum filtration, a variation of generalized Wiener filter can be used. This approach is illustrated by Figures 3,4 for a simplified case of obtaining of 12 dB loop attenuation, uniform across 4000Hz frequency range.

![Figure 2.1. The short-term spectrum (linear domain) of RCV and SND signals before post-processing. Note that some of formants coincide in frequency domain, but the most are located in different spots.](image1)

In a simplified approach, an AEC should insert 12 dB in either of the direction, or split 6 dB and 6 dB for each channel. The GAEC’s post-processor builds sub-band filters in both directions so that the major formants of both voice signals are preserved as much as possible in a given situation. If both near-end and far-end talkers happen to pronounce similar phonemes at the same time, this approach bears no advantage over the basic AEC. In all other cases this approach provides superior intelligibility for both RCV and SND/DT signals. This approach is most effective on moderate attenuation levels.

![Figure 2.3. The short-term spectrum (linear domain) of RCV and SND signals after Wiener-filter-based post-processing. The loop attenuation (both RCV and SND) curve is in green (0.25 = 12 dB in this particular example). Note that although the shape of the short-term spectrum curves is altered in areas of significant overlap, the placement and presence of formants is preserved.](image2)

The GAEC post-filter is regularized to keep balance between sharp formant filtering, maximizing of the residual echo removal, voice dynamics (10s of ms), the room echo path decay properties, and excessive modification of the voice signal, what may result in a “robotic” sound. The further details of the actual implementation are beyond the scope of this document.
2.1.3 Noise reduction

The noise reduction algorithm is loosely based on the industry-adopted Ephraim-Malah noise reduction approach, which is both optimal in a minimal mean-square error sense and well known for its associated colorless residual noise.

By definition, noise is a random signal, and therefore its exact value at any given moment of time cannot be predicted. Yet, if after observing a random signal for a period of time we can conclude that it consistently displays nearly stationary behavior, we can attempt to qualify incoming signal to be either noise or not. The qualification is not binary, and some attenuation is inserted accordingly to its results and the noise distribution curve.

This approach is capable of significant reduction of the noise level if and only if the noise is stationary and it is of Gaussian nature, as it is in the case of thermal circuit noise. Generally, this is not the case with typical office noise, which contains both stationary components (as the noise of computers, air conditioners, fans, thermal noise of the microphone circuitry) and non-stationary burst sources as sounds of keyboard typing, humans walking around, talking in the background, etc.

Typically, the level of office noise picked up by a speakerphone’s microphone is reduced by 6-15 dB. The most annoying monotonic noise sources as fan noise is reduced the most, but little reduction can be expected for the non-stationary noise sources. Users tend to consider the presence of such bursts positively as conveying the feeling of staying connected with far-end (opposite to half-duplex mode).

GAEC performs noise reduction accordingly to the regularized MIPS optimized MMSE variation of the Ephraim-Malah approach. The further details of the actual implementation are beyond the scope of this document.
2.2 Processing blocks

Figure 2.4. The GAEC functional block diagram. Receive path processing blocks are in cyan, Send – red, Adaptation – green, and Control – yellow.

2.2.1 Receive path processing

RCV Direct Analysis Filterbank performs sub-band complex domain decomposition of the input RCV signal into sub-bands. RCV Post-processing Filter processes the RCV signal, when and how required by the control block.

Perceptually Weighted Noise Injection block inserts noise in the RCV signal to improve the eigen-values distribution. That improves, in its turn, the adaptive processing convergence. The possibility of insertion of quite high noise in critical bands without human ear noticing is possible due to the frequency masking effect, one of the basic principles of so-called MP3 codecs. The configuration parameter pCfg->sNoise determines how much noise is inserted, to enable field adjustments. The noise is injected only when required.

RCV Direct Synthesis Filterbank block forms the RCV path output signal out of the sub-band decomposed signal to go to the loudspeaker.
2.2.2 Send Path processing
SND Analysis Filterbank block performs sub-band complex domain decomposition of the input SND signal into sub-bands. In the Sub-Band Echo Cancellation block, the best estimate of the echo is subtracted from the sub-band RCV. SND Post-processing Filter operates in conjunction with RCV Post-processing Filter. Noise Reduction block reduces the level of stationary noise. SND Synthesis Filterbank block forms the SND path output signal out of the sub-band decomposed signal to go to the far-end party.

2.2.3 Adaptation
RCV Analysis Filterbank decomposes RCV OUT signal into sub-bands for echo cancellation. Sub-band Adaptation block operates with two filters: an active filter and its “mirror” image, which hold the best stable estimate of the echo path. The echo tail saturation is performed on each step to control situations when a near-end user unintentionally closes loudspeaker-microphone acoustic path with a palm.

Adaptation Speed Control block performs step size (Calman gain) control, accounting for loudspeaker’s non-linearity, adaptively learnt SND noise level, and the echo path variation curves, and decides when the mirror filter can be updated. GAEC uses 3rd order model for loudspeaker distortion approximations.

2.2.4 GAEC control
GAEC 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 whenever the RCV IN signal spectrum contains isolated peaks.

NOTE: If DTMF tones are clipped because they have been amplified, they may not be recognized as pure sine waves. GAEC shall be notified explicitly via control interface in this case.

Double Talk Detector (DTD) determines if any sub-band (or entire band) experience double talk, and commands appropriate actions to all other blocks. Echo Path Change Detector decides if echo path has significantly changed and re-converging is necessary. Non-Linear Post-processor Control (NLP) provides control over Post-Processing Filters to ensure appropriateness of their actions.

Signal Parameter Assessment provides per-band signal measurements for all other blocks, and shapes the measurement result in a special functional uniform representation. Most of these control blocks operate with variables rather than with discrete states.

General GAEC Control governs over GAEC control flow and signal flow, along with checking for sanity of GAEC operations and collecting statistics.

Further details are beyond the scope of this document.
3 API

3.1 Data structures

3.1.1 Configuration

Configuration is an object for passing of control information to a GAEC instance. The data is copied into GAEC database during initialization or a control call. Note that `Int` is defined in TI’s supplied `ialg.h` file as 16 bit signed word.

```c
typedef struct IGAEC_tCfg {
  Int sErlMin;
  Int sDtThr;
  Int sRcvDistThr;
  Int sTCLst;
  Int sTCLdt;
  Int sWhiteThr;
} IGAEC_tCfg;
```

The parameters are expressed in 0.1 dB units (3.5 Db is expressed as 35).

The fields of configuration have the following meaning:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Recommended Range</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>ErlMin</td>
<td>Minimum ERL of the echo path. The value is used during initial convergence only.</td>
<td>2...8 dB</td>
<td>6 dB</td>
</tr>
<tr>
<td>DtThr</td>
<td>A DT signal shall exceed the expected echo level for more than DtThr to be accounted for. Higher setting makes GAEC more robust to false DT detection, but shrinks DT range.</td>
<td>2...6 dB</td>
<td>3 dB</td>
</tr>
<tr>
<td>RcvDistThr</td>
<td>The level of wide band signal (in dBm0) that causes approximately 1% of inter-modulation distortion of the loudspeaker.</td>
<td>-</td>
<td>-18.0 dB</td>
</tr>
<tr>
<td>TCLst</td>
<td>The target per-band Terminal Coupling Loss for the single talk originated on the far-end side.</td>
<td>25...55 dB</td>
<td>40 dB</td>
</tr>
<tr>
<td>TCLdt</td>
<td>The least per-band TCL for the double talk condition when residual echo is partially masked by near-end signal.</td>
<td>10...20 dB</td>
<td>15 dB</td>
</tr>
<tr>
<td>WhiteThr</td>
<td>This threshold controls amount of noise being added to RCV OUT signal.</td>
<td>-25...-17 dB</td>
<td>-20dB</td>
</tr>
</tbody>
</table>

Both TCLst and TCLdt settings depend on the delay between near- and far-ends, and expected far-end sidetone level. The recommended values are (which might be relaxed in some cases):

<table>
<thead>
<tr>
<th>The type of Connection</th>
<th>TCLst</th>
<th>TCLdt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calls within VoLAN (4-wire) systems with less than 20 ms of one-way delay.</td>
<td>30</td>
<td>10...15</td>
</tr>
<tr>
<td>External local calls from VoLAN (4-wire)</td>
<td>35...40</td>
<td>15</td>
</tr>
<tr>
<td>Long distance calls over PSTN, if interactions with a network EC may occur.</td>
<td>45</td>
<td>20</td>
</tr>
<tr>
<td>WAN VoIP calls (unpredictable delays)</td>
<td>55...60</td>
<td>30...40</td>
</tr>
</tbody>
</table>

There is no need to assign high values to the TCLst / TCLdt because attenuation is performed in quite narrow sub-bands with the width close to the critical band’s width.\(^{15}\) Do not configure GAEC for “the-worst-case” of WAN VoIP if the call is indeed VoLAN / local unless half-duplex behavior is desirable.

The echo tail length is not a configuration, but a build parameter. The structure of GAEC is designed to be efficient for echo tails less than 400...500 ms. The length of echo tail and the maximum achievable

\(^{15}\) ITU-T Recommendation P.342 'Transmission characteristics for telephone band (300-3400 Hz) digital loudspeaking and hands-free telephony terminals' (05/2000) suggests setting of in-band TCL about 10 dB lower than TCLw, if the bands are 1/3 octave wide.
TCLwst shall be balanced. The most appropriate library for the given speakerphone hardware is supplied. Multiple libraries can be provided upon request.

GAEC performance may degrade or GAEC may become dysfunctional if the configuration parameters are inappropriate. Ensure that a user understands the meaning of parameters and the consequences of wrong settings before any changes are applied.

### 3.1.2 GAEC statistics

GAEC provides the following statistics:

```c
typedef struct GAEC_MIKET_tStts {
  int sRcvNoise;
  int sSndNoise;
  int sRcvAtt;
  int sSndAtt;
  int sErl;
  int sErle;
  int sMaxCoef;
  int sWorstErl;
} GAEC_MIKET_tStts;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RcvNoise</td>
<td>RCV IN noise level in dBm0</td>
</tr>
<tr>
<td>SndNoise</td>
<td>SND IN noise level in dBm0</td>
</tr>
<tr>
<td>RcvAtt</td>
<td>Weighted average RCV path attenuation</td>
</tr>
<tr>
<td>SndAtt</td>
<td>Weighted average AND path attenuation</td>
</tr>
<tr>
<td>Erl</td>
<td>Average ERL</td>
</tr>
<tr>
<td>Erle</td>
<td>Average ERLE</td>
</tr>
<tr>
<td>MaxCoef</td>
<td>The value reported shall be less than 25,000. See ‘Integrating GAEC…” for more details</td>
</tr>
<tr>
<td>Worst Erl</td>
<td>The value reported shall be less than sMinErl configuration parameter. It is highly desirable that it is less than 0 dB. See ‘Integrating GAEC” for more details</td>
</tr>
</tbody>
</table>

Energies are expressed in 0.1 dB units with precision of approximately 0.25 dB.

### 3.2 IALG interface

TI’s “Express DSP” compliant interface is fully supported. See relevant TI documentation for further details. Memory allocation details for non-initialized segments (see ‘gaec_miket.h’ for the latest data):

<table>
<thead>
<tr>
<th>No</th>
<th>Element</th>
<th>Referred as</th>
<th>Length (words) for 0 echo tail</th>
<th>Increment (words) per 100 ms of echo tail</th>
<th>Alignment (words)</th>
<th>Preferable allocation in</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>xDAIS Obj</td>
<td>-</td>
<td>12</td>
<td></td>
<td>2</td>
<td>SARAM/extend</td>
</tr>
<tr>
<td>2</td>
<td>Core</td>
<td>pDb, .gaecdb</td>
<td>1358</td>
<td>240</td>
<td>2</td>
<td>SARAM</td>
</tr>
<tr>
<td>3</td>
<td>Signal History</td>
<td>pHst, .gaech</td>
<td>60</td>
<td>1600</td>
<td>2</td>
<td>DARAM</td>
</tr>
<tr>
<td>4</td>
<td>Adaptive Filter</td>
<td>pAdf, .gaecaf</td>
<td>0</td>
<td>1600</td>
<td>2</td>
<td>SARAM</td>
</tr>
<tr>
<td>5</td>
<td>Mirror of Adaptive Filter</td>
<td>pAdfM, .gaecafm</td>
<td>0</td>
<td>1600</td>
<td>2</td>
<td>SARAM</td>
</tr>
<tr>
<td>6</td>
<td>Scratch-pad</td>
<td>pSc, .gaecsc</td>
<td>646</td>
<td></td>
<td>2</td>
<td>DARAM</td>
</tr>
</tbody>
</table>

Scratch pad is used during GAEC_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 GAEC_MIKET_process(…).
Ideally, each element shall reside in a different memory segment. Performance will suffer if the database elements (especially signal history and adaptive filter arrays) are placed in external RAM. Refer to “Integrating GAEC into a Speakerphone” chapter for further details.

3.3 Vendor specific API

3.3.1 Initialization

```c
extern void GAEC_MIKET_init_db(
    void *pDb,
    IGAEC_tCfg *pCfg,
    Int *pHst,
    Int *pAdf,
    Int *pAdfM);
```

pCfg shall refer to a valid configuration structure; pDb, pHst, pAdf, pAdfM shall refer to user-allocated long word aligned memory blocks.

3.3.2 Control

```c
typedef struct IGAEC_Status {
    Int size;           /* sizeof the whole parameter struct */
    IGAEC_tCfg *pCfg;   /* ptr to new cfg, if _CMD_CFG set */
    Int sVolume;        /* Rcv volume in 0.1dB units, if CMD_VOLUME*/
    Int uCmdMask;  /* what is affected */
} IGAEC_Status;
```

```c
extern void GAEC_MIKET_control(
    void *pDb,
    Int Cmd,
    IGAEC_Status *pStatus);
```

<table>
<thead>
<tr>
<th>Command</th>
<th>Value</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGAEC_CMD_RESET</td>
<td>0x0001</td>
<td>Reset GAEC</td>
</tr>
<tr>
<td>IGAEC_CMD_PRESERVE_ADF</td>
<td>0x0002</td>
<td>During reset, preserve adaptive filter from zeroing, to use the same filter for next call</td>
</tr>
<tr>
<td>IGAEC_CMD_VOLUME</td>
<td>0x0004</td>
<td>Apply new volume</td>
</tr>
<tr>
<td>IGAEC_CMD_CFG</td>
<td>0x0008</td>
<td>Replace configuration</td>
</tr>
<tr>
<td>IGAEC_CMD_ALL_OFF</td>
<td>0x0010</td>
<td>Disable and bypass</td>
</tr>
<tr>
<td>IGAEC_CMD_ADAPT_OFF</td>
<td>0x0020</td>
<td>Disable adaptation</td>
</tr>
<tr>
<td>IGAEC_CMD_CNL_OFF</td>
<td>0x0040</td>
<td>Disable echo cancellation (TBD)</td>
</tr>
<tr>
<td>IGAEC_CMD_RCV_NSE_OFF</td>
<td>0x0080</td>
<td>Disable additive noise on RCV path</td>
</tr>
<tr>
<td>IGAEC_CMD_NLP_OFF</td>
<td>0x0100</td>
<td>Disable NLP post-processing</td>
</tr>
<tr>
<td>IGAEC_CMD_NSE_RED_OFF</td>
<td>0x0200</td>
<td>Disable noise reduction</td>
</tr>
<tr>
<td>IGAEC_CMD_SND_MUTED</td>
<td>0x1000</td>
<td>Informs GAEC than SND mute is applied</td>
</tr>
<tr>
<td>IGAEC_CMD_LOOPBACK</td>
<td>0x2000</td>
<td>Enforces self-testing loopback</td>
</tr>
<tr>
<td>IGAEC_CMD_TONE</td>
<td>0x4000</td>
<td>Informs GAEC that a tone is active on RCV IN.</td>
</tr>
</tbody>
</table>

Cmd can be constructed by OR-ing flags. The _CFG, _RESET, and _VOLUME actions do not require command masking. In all other cases, the pStatus->uCmdMask contains bit mask for all commands applied. Set regarding Cmd bit to activate an action (otherwise it will be deactivated).

It is not recommended to apply _ALL_OFF without _RESET (with or without _PRESERVE_ADF) due to the obvious consequences of re-enabling GAEC by calling _control(pDb, 0); afterwards.

Asynchronous calling _control() and _process() from different tasks is not recommended.
3.3.3 Process

extern void GAEC_MI KET_process(
    void *pDb,
    void *pSc,
    Int *pRcv,
    Int *pSnd);

- pDb shall refer to the initialized core.
- pSc shall refer to scratch pad.
- pRcv / pSnd shall point to least significant bit (4096 as 0 dBm RMS) aligned frames of 40 sample continuous data.
4 Integrating GAEC into a speakerphone

4.1 Interfacing and distortion correction filters

![Diagram of GAEC's integration into packet voice processing solutions.](image)

Figure 4.1. The diagram of GAEC’s integration into packet voice processing solutions.

There shall be absolutely no ‘non-linear’ signal alternations between RCV OUT and the loudspeaker, as well as between microphone input and pre-SND correction filter.

GAEC shall be explicitly notified of the Mute operation, and the actual Mute shall be performed after GAEC signal processing. GAEC will provide no attenuation on the RCV path because people often use MUTE button to ensure they hear everything what far-end is saying.

GAEC shall be explicitly notified if the volume settings on the speakerphone are modified, because GAEC keeps track of the RCV noise level. If the noise level suddenly jumps up or down, GAEC may react inappropriately till it adapts to new noise level.

The variance of loudspeaker and microphone sensitivity shall be as low as possible, and be within +/- 2dB to achieve repeatable GAEC performance.

The 3 equalizing filters on the diagram shall be designed specifically for each customer’s speakerphone cabinet and hardware. The performance of speakerphone with GAEC shall be evaluated using a controllable environment, and it shall be ensured that overall transfer functions are sufficiently flat, learnt loudspeaker’s non-linearity is soft, and the cross-talk ERL is reasonable. The sMaxCoef parameter, reported by statistics, shall have a satisfactory value, as well as the sWorstErl. Moreover, the equalizing filters may impact perceptual loudness and, therefore, TCLwst/TCLwtd. Default configuration shall be adjusted accordingly.

Generally, integration of any AEC into a speakerphone represents a complex task, which shall be performed by a knowledgeable and experienced party. After integration, the speakerphone as a whole shall be tested for conformance with ITU-T Recommendations to the extent determined by a customer:

- P.330 ‘Speech Processing Devices for Acoustic Enhancement’ (03/2003),
- P.340 ‘Transmission characteristics and speech quality parameters of hands-free terminals’ (05/2000).
- P.342 ‘Transmission characteristics for telephone band (300-3400 Hz) digital loudspeaking and hands-free telephony terminals’ (05/2000).

GAEC may need to be modified for a particular speakerphone to workaround its limitations and improve overall performance.
### 4.2 MIPS and Memory

The following data is provided to facilitate in choosing the most appropriate DSP part. The MIPS tables are approximate and provided as a guideline.

#### 4.2.1 Memory consumption.

Refer the map file (obtained during GAEC library linkage) for latest data.

<table>
<thead>
<tr>
<th>Section name</th>
<th>Page</th>
<th>Initialized</th>
<th>Size</th>
<th>Increase per 100 ms of echo tail</th>
<th>Priority of placement in internal RAM</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAGE 0, <code>C54x</code></td>
<td>6454</td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.text</td>
<td>0</td>
<td>Y</td>
<td>2594</td>
<td>0</td>
<td>6 (lowest)</td>
<td>Slow “C”, mostly control code</td>
</tr>
<tr>
<td>.ftext*</td>
<td>0</td>
<td>Y</td>
<td>3460</td>
<td>&lt;3</td>
<td>4</td>
<td>Fast assembly code, can change slightly with echo tail change</td>
</tr>
<tr>
<td>.smcode</td>
<td>0</td>
<td>Y</td>
<td>400</td>
<td>0</td>
<td>1 (highest)</td>
<td>Self-Modifying Code of tight adaptation loops themselves (quasi-floating point)</td>
</tr>
<tr>
<td>PAGE 0, <code>C55x</code></td>
<td>6187</td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.text</td>
<td>0</td>
<td>Y</td>
<td>2688</td>
<td>&lt;2</td>
<td>6</td>
<td>Slow “C”, mostly control code</td>
</tr>
<tr>
<td>.ftext2</td>
<td>0</td>
<td>Y</td>
<td>1871</td>
<td>&lt;3</td>
<td>5</td>
<td>Moderately fast assembly code, the code size can change due to constant’s sizing</td>
</tr>
<tr>
<td>.ftext1</td>
<td>0</td>
<td>Y</td>
<td>1628</td>
<td>&lt;3</td>
<td>4</td>
<td>Fast assembly code, the code size can change due to constant’s sizing</td>
</tr>
<tr>
<td>PAGE 1</td>
<td>15692</td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.gaectab</td>
<td>1</td>
<td>Y</td>
<td>3900</td>
<td>0</td>
<td>5</td>
<td>Tables</td>
</tr>
<tr>
<td>.gaecsc</td>
<td>1</td>
<td>N</td>
<td>646</td>
<td>0</td>
<td>1</td>
<td>Scratch pad</td>
</tr>
<tr>
<td>.gaecaf</td>
<td>1</td>
<td>N</td>
<td>3200</td>
<td>1600</td>
<td>3</td>
<td>Adaptive filter</td>
</tr>
<tr>
<td>.gaeafm</td>
<td>1</td>
<td>N</td>
<td>3200</td>
<td>1600</td>
<td>5</td>
<td>Adaptive filter mirror</td>
</tr>
<tr>
<td>.gaech</td>
<td>1</td>
<td>N</td>
<td>3260</td>
<td>1600</td>
<td>2</td>
<td>Signal history</td>
</tr>
<tr>
<td>.gaeedb</td>
<td>1</td>
<td>N</td>
<td>1486</td>
<td>240</td>
<td>5</td>
<td>Data base</td>
</tr>
<tr>
<td>Total, <code>C54x</code>, for 200ms echo tail</td>
<td>22144</td>
<td>5040</td>
<td>Total number for <code>C55x</code> is 267 words less</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Initialized, <code>C54x</code></td>
<td>10356</td>
<td>&lt;3</td>
<td>.&quot;.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Non-Initialized, <code>C54x</code></td>
<td>11790</td>
<td>5040</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Standard sections**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
<th>Size</th>
<th>Priority</th>
</tr>
</thead>
<tbody>
<tr>
<td>.stack</td>
<td>1</td>
<td>&lt;50</td>
<td>6</td>
</tr>
<tr>
<td>.bss</td>
<td>1</td>
<td>0</td>
<td>Not used</td>
</tr>
<tr>
<td>.data</td>
<td>1</td>
<td>0</td>
<td>Not used</td>
</tr>
<tr>
<td>.switch</td>
<td>0</td>
<td>0</td>
<td>Not used</td>
</tr>
<tr>
<td>.const</td>
<td>1</td>
<td>0</td>
<td>Not used</td>
</tr>
</tbody>
</table>
4.2.2 MIPS on C54x

The MIPS usage heavily depends on the sections’ placement in memory map. The following table represents results of GAEC profiling for the case of 240 ms echo tail on C5402 DSK (100 MHz, 1 wait state external RAM). All other sections are in external RAM.

- DARAM0 is referred as D0
- DARAM1 is referred as D1
- External RAM is referred as X.

<table>
<thead>
<tr>
<th>.text</th>
<th>.gaecdb</th>
<th>.gaecaf</th>
<th>.gaecafm</th>
<th>.gaech</th>
<th>MIPS for 0 ms</th>
<th>MIPS increase /100 ms</th>
<th>Internal RAM usage for 0 ms</th>
<th>Internal RAM usage increase /100 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>&gt;100</td>
<td>744</td>
<td>400</td>
<td>23116</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>D0</td>
<td>&gt;100</td>
<td>4644</td>
<td>400</td>
<td>19216</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>D1</td>
<td>X</td>
<td>D0</td>
<td>86.21</td>
<td>4644</td>
<td>4240</td>
<td>15376</td>
</tr>
<tr>
<td>D1</td>
<td>X</td>
<td>D1</td>
<td>X</td>
<td>D0</td>
<td>71.23</td>
<td>4644</td>
<td>7700</td>
<td>11916</td>
</tr>
<tr>
<td>D1</td>
<td>D0</td>
<td>D1</td>
<td>X</td>
<td>D0</td>
<td>66.30</td>
<td>6226</td>
<td>7700</td>
<td>10334</td>
</tr>
</tbody>
</table>

The MIPS consumption for other echo tail lengths can be estimated with the following table:

<table>
<thead>
<tr>
<th>.text</th>
<th>.gaecdb</th>
<th>.gaecaf</th>
<th>.gaecafm</th>
<th>.gaech</th>
<th>MIPS for 0 ms</th>
<th>MIPS increase /100 ms</th>
<th>Internal RAM usage for 0 ms</th>
<th>Internal RAM usage increase /100 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>40</td>
<td>45</td>
<td>684</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>D0</td>
<td>40</td>
<td>27</td>
<td>744</td>
<td>1600</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>D1</td>
<td>X</td>
<td>D0</td>
<td>40</td>
<td>18</td>
<td>744</td>
<td>3200</td>
</tr>
<tr>
<td>D1</td>
<td>X</td>
<td>D1</td>
<td>X</td>
<td>D0</td>
<td>25</td>
<td>18</td>
<td>4204</td>
<td>3200</td>
</tr>
<tr>
<td>D1</td>
<td>D1</td>
<td>D0</td>
<td>D0</td>
<td>D0</td>
<td>23</td>
<td>15</td>
<td>4204</td>
<td>4800</td>
</tr>
<tr>
<td>D1</td>
<td>D0</td>
<td>D1</td>
<td>D0</td>
<td>D0</td>
<td>20</td>
<td>14</td>
<td>5210</td>
<td>5040</td>
</tr>
</tbody>
</table>

4.2.3 MIPS on C55x

The MIPS were measured on ‘C5510 DSK with 100 MHz external SDRAM. Scratch pad, stack, .gaecaf, .gaech have been placed in internal RAM (DARAM) during this test; .gaectab into SARAM.

<table>
<thead>
<tr>
<th>.gaecafm</th>
<th>.text</th>
<th>.ftext2</th>
<th>.ftext1</th>
<th>.gaecdb</th>
<th>MIPS for 0 ms</th>
<th>MIPS increase /100 ms</th>
<th>Internal RAM usage for 0 ms</th>
<th>Internal RAM usage increase /100 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>DARAM</td>
<td>SARAM</td>
<td>SARAM</td>
<td>SARAM</td>
<td>DARAM</td>
<td>14</td>
<td>9</td>
<td>11781</td>
<td>5040</td>
</tr>
<tr>
<td>X</td>
<td>SARAM</td>
<td>SARAM</td>
<td>SARAM</td>
<td>DARAM</td>
<td>14.5</td>
<td>12</td>
<td>11781</td>
<td>3440</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>SARAM</td>
<td>SARAM</td>
<td>SARAM</td>
<td>19.5</td>
<td>12</td>
<td>9093</td>
<td>3440</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>SARAM</td>
<td>DARAM</td>
<td>27.5</td>
<td>12</td>
<td>7222</td>
<td>3440</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>DARAM</td>
<td>39</td>
<td>10</td>
<td>5594</td>
<td>3440</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>50</td>
<td>12</td>
<td>4606</td>
<td>3200</td>
</tr>
</tbody>
</table>

A specific combination of the echo tail length with a particular device of the ‘C55x family may require specific, unique segment placement.

Be strongly advised against placing sections with intensively used data (as .gaech) into external SDRAM.
5 Specifications

The specification data below follows ITU-T Recommendation G.167 ‘Acoustic Echo Controllers’. Familiarity with this and other related recommendations is beneficial for readers desiring achieve detailed understanding of the specifications given below.

The performance of a speakerphone depends both on the quality of the AEC solution it uses and the acoustic qualities of its loudspeaker, microphone and the housing. To decouple the GAEC specifications from the limitations and influences imposed by external to GAEC speakerphone components, GAEC is first characterized by off-line bit-exact simulation assuming fully linear electro-acoustical components.

5.1 Transmission specifications

- GAEC is a narrow-band (8 kHz sampling rate) echo control component.
- GAEC does not modify the bandwidth of the terminal within 300…3400 Hz frequency mask.
- When in single talk mode, GAEC contribute less than ±1dB to the total attenuation distortions within the 300…3400 Hz bandwidth.
- GAEC inserts 6.25 ms of delay in both RCV and SND paths. It raises a problem when GAEC is used in 2-wire phones, which is resolved with line EC16.
- When in single talk mode, GAEC contributes less than 1ms to the delay distortions.
- The non-linear distortions, added by GAEC due to its 16-bit fixed-point implementation, are low.
- GAEC does not emit noise on the send side.
- Subject to correctness of the pCfg->sWhiteThr configuration parameter setting, GAEC does not produce a subjectively noticeable disturbance on receive side.

5.2 Acoustic echo control specifications

5.2.1 Measurement signal

Stationary noise with average spectrum complying with Recommendation P.50 is used17. TCLw measurements are performed as specified by Recommendation G.12218. The GAEC’s performance on real-life voice signal is discussed in the chapters 6 and 7.

5.2.2 Acoustic echo path

Three distinct real acoustic echo paths were measured using a sample speakerphone supplied by a customer:

- A small (~15 m³) acoustically treated room with reverberation time of about 80 ms.
  - 400 Hz: $T_{60} = 140$ ms
  - 2000 Hz: $T_{60} = 63$ ms
  - 3200 Hz: $T_{60} = 88$ ms
- A medium-to-large (~100 m³) office, often used as a board room, with reverberation time of about 520 ms.
  - 400 Hz: $T_{60} = 660$ ms
  - 2000 Hz: $T_{60} = 495$ ms
  - 3200 Hz: $T_{60} = 440$ ms
- A large warehouse 2-story room (~500 m³) with reverberation time of about 1.1 s.
  - 400 Hz: $T_{60} = 1280$ ms
  - 2000 Hz: $T_{60} = 990$ ms
  - 3200 Hz: $T_{60} = 1090$ ms

16 MIKET DSP Solutions can provide a high-quality line echo canceller (RC-RLS).
17 GAEC was not optimized to the spectral properties of this noise signal. Moreover, GAEC is incapable of using of some of its properties to achieve its best performance on noise.
The speakerphone used for measurement was not designed with an AEC concerns taken into consideration. It has severe distortion on both receive and send path, and high loudspeaker-microphone coupling via multiple standing waves, which fall to –60 dB level within approximately 50 ms. The equalization filters need to be applied to make the transfer functions smoother. The internal coupling represents the major part of the echo for all rooms.

![Figure 5.1. The echo density of the first 300 ms of the measured echo path used for GAEC simulation. Note that the first 15 ms of the echo represent the internal to speakerphone resonance modes.](image1)

![Figure 5.2. The transfer functions of 3 real echo paths display negative ERL for half of the frequency range; therefore, an acoustic echo of a valid RCV signal can easily saturate SND codec. Appropriate hardware and software precautions must be taken in such situations.](image2)

To create an additional set of echo paths for the simulation of a speakerphone with a good acoustic design, the initial 15 ms of all echo paths were attenuated by 20 dB. Most of the rooms and speakerphones shall fall within margins set by those six acoustic echo paths.
The TCL\textsubscript{wst} for those echo paths (measured with adaptation disabled) are:
- $T_{60} = 80\text{ms}: 1.75\text{ dB. Modified variation: 18.97 dB}$
- $T_{60} = 520\text{ms}: 1.43\text{ dB. Modified variation: 13.53 dB}$
- $T_{60} = 1100\text{ms}: 0.43\text{ dB. Modified variation: 14.84 dB}$

Modified variations of paths are indicated by ‘M’ letter (‘M80ms’, ‘M520ms’, ‘M1100ms’) in the rest of this chapter.

### 5.2.3 Weighted terminal coupling loss – single talk (TCL\textsubscript{wst})

Requirement: TCL\textsubscript{wst} shall be at least 40 dB.

TCL\textsubscript{wst} is a user-configurable parameter (pCfg->sTCLst).

### 5.2.4 Weighted terminal coupling loss – double talk (TCL\textsubscript{wdt})

Requirement: TCL\textsubscript{wdt} shall be at least 25 dB.

TCL\textsubscript{wdt} is a user-configurable parameter (pCfg->sTCLdt).
There is no need in the measurements. Moreover, the full-band approach of G.167 is not directly applicable to GAEC because GAEC performs weighting in narrow sub-bands.

5.2.5 Received and sent speech attenuation during double talk (Ardt/Asdt)

Requirement: Ardt/Asdt shall be no more than 6dB.

Ardt/Asdt is computed by GAEC as a half of the difference between pCfg->sTCLdt and achieved (ERL+ERLE = adaptive part of TCLwst) on per-band basis (or 0, if ERLE + ERL > TCLdt). Per-band Ardt/Asdt is less than 6 dB if that difference is less than 12 dB. Typically, Ardt/Asdt are 0.

5.2.6 Received and sent speech distortion during double talk (Drdt/Dsdt)

G.167 does not specify any particular measures for these distortions. The nature of these distortions due to post-processing in GAEC has been explained in the ‘Theory of Operations’ chapter. The detailed discussing of the voice distortions is given in the chapter 6.

5.2.7 Maximum frequency shift

Requirement: under study.

GAEC does insert frequency shifts because hands-free terminals are likely to be used on connections including network electric echo cancellers conforming to ITU-T G.168 ‘Digital network echo cancellers’ which are not able to work properly with time-variant echo paths (e.g. frequency shift).

5.2.8 Break-in time – single talk (Tonst)

Requirement: Tonst_r / Tonst_s shall be no more than 20ms.

The GAEC structure is more complicated than G.167 accounts for. Simple measuring attenuation of the RCV / SND signal in wide-band energy terms is not applicable. The onset times can be estimated only from observing input and output signals. The time scales are merged on RCV OUT.

Figure 5.5. Transition from RCV to SND activity. SND IN signal starts at 5007 ms, and the SND OUT signal follow the input very closely from its onset (difference in energy is less than 3 dB). High frequency components are more attenuated than low frequency ones.
This test was undertaken for echo path of 520ms, with echo tail of 250ms. There is no significant difference between echo paths for GAEC on transition between RCV and SND activity. Real voice is easier for switching than noise bursts with amplitude changing momentarily from 0 to maximum and back because a human throat is not capable of momentarily developing or squelching of strong vowels.

5.2.9 Break-in time – double talk (Tondt)

Requirement: The attenuation of RCV OUT / SND OUT signals shall be no more than 6dB after 20 ms.

Refer to chapter 6 for detailed discussion of GAEC double-talk performance on voice signals.

5.2.10 Initial convergence time (Tic)

Requirement: the attenuation of the echo shall be at least 20 dB after 1s.

Convergence starts after GAEC has acquired RCV IN signal for the duration of the echo tail the GAEC has been built for. Then GAEC converges to the echo path, starting with a fast RLS-like mode and gradually switching to the Calman mode. The convergence speed depends on the echo tail length (in inverse proportion), and on the energy of the real echo beyond GAEC’s echo tail capability. As a rule of thumb, gaining of each 3 dB requires doubling of the convergence time, in strict accordance with asymptotic theory ($\sigma = \frac{1}{\sqrt{t}}$).
The initial speed of convergence for modified echo paths with higher echo return loss (ERL) is lower, but the curves tend to merge later on - in full accordance with the theory of adaptive signal processing.

5.2.11 Recovery time after double talk (Trdt)

Requirement: the attenuation of the SND signal shall be at least 20 dB after 1s (after end of double talk).

GAEC employs a sensitive double talk detector, and only a very low-level double talk can pass undetected. If the double talk is detected, no degradation in quality happens, and recovery time is zero. If double talk passes undetected, the TCLwst recovery follows the same convergence curve as in the previous test.

5.2.12 Terminal coupling loss during echo path variations (TCLwpv)

Requirement: TCLwpv shall be at least 10 dB.

No particular means to produce echo path variations are given in G.167, nor size of those variations is discussed.

GAEC processes echo path variations accordingly to their severity.

- Small variations: There are always some variations in echo path due to people breathing, talking, slightly moving their heads, hands, etc, and these often lead to the limiting TCLwst to the 30…35 dB. GAEC adapts to these small variations by using Calman filtration.
- Large variations: When people move in speakerphone’s proximity, a door opens, or someone adjusts volume on the speakerphone with a palm over its elements, the echo path undergoes significant changes. GAEC recognizes this situation with an echo path change detector, which is human voice specific. GAEC displays inferior performance when being tested with noise signal.

To simulate a strong but possibly realistic echo path change, the coefficients of echo path (for an office room with $T_{60}$ of 520ms) beyond 75 ms are all inverted at the 5s time point. The initial part of the echo path, responding for speakerphone’s internal resonance modes, remains unchanged.
Figure 5.9. The SND OUT signal jumps at 5s because GAEC has no means to distinguish between echo path change and double talk. Although adaptive part of TCL (green) falls below 10 dB, the cumulative TCLwst is kept above 10 dB by attenuating RCV OUT signal. The RCV OUT signal falls below –30 dBm (RCV IN level less 20 dB) in 400ms. Full recovery follows in approximately 4…5 seconds.

5.2.13 Recovery time after echo path variations (Trpv)
Requirement: the attenuation of the echo shall be at least 20 dB after 1s.

See Figure 5.8 in the previous chapter.

5.3 **Specification for interworking with the network**

5.3.1 **Interworking with speech codecs**
All mentioned considerations were accounted for during GAEC design.

5.3.2 **Interworking with network echo cancellers**
All mentioned considerations were accounted for during GAEC design. GAEC starts with operating in pure half-duplex mode, with switching of maximal attenuation, to prevent false convergence of network echo cancellers. Subject to the correctness of the pCFG->sTCLst threshold, interworking with network echo cancellers, conforming to ITU-T G.168, shall be satisfactory.

5.3.3 **DCME and PCME interworking**
All mentioned considerations were accounted for during GAEC design. GAEC keeps the background noise level on SND OUT as low as possible, mixing in comfort noise with matching level and spectrum during SND post-processing. Subject to correctness of pCfg->sTCLst setting, background noise may have inserts of residual echo on high-level RCV signal.

Note that the background noise in a typical office, after being picked up by a microphone with a high gain amplifier attached, does not tend to be of a constant level, and this is out of the GAEC control.

5.3.4 **Interworking between a wide-band terminal and other types of terminals through the network.**
Not applicable. GAEC is a narrow band AEC.
5.4 Additional specifications

5.4.1 Convergence for various RCV IN levels

The adaptive part of the TCLwst value depends on the RCV IN signal. GAEC is capable of decreasing the residual error down to –65…-70 dBm, if the speakerphone’s acoustic is good. Note that for office rooms the level of background noise usually falls in the –50…60 dBm range.

![Figure 5.10. GAEC convergence curves for various levels of RCV IN. Echo tail size is 250 ms.](image)

5.4.2 Double talk sensitivity

To characterize GAEC double talk sensitivity, a periodic 800 Hz square wave signal is used. The spectrum of this square wave has odd harmonics at 800, 2400 and 4000 Hz, what is distinctively different from the RCV IN noise (which has solid spectrum with –6 dB slope after 500 Hz). This kind of discrepancy is quite typical for customary double-talk situations because near- and far-end talkers would rarely pronounce the same vowel with the same pitch.

Test conditions:
- The echo path with T60 of 520ms is used.
- GAEC’s echo tail is 250 ms.
- RCV IN signal has –10dBm intensity.
- The adaptive part of TCLwst by ‘0’ time is about 31 dB. Residual echo has level of –45 dB (23 units RMS) before post-processing and noise reduction.

![Figure 5.11. SND OUT signal with input double talk signal on –20 dBm (410 in units) passes almost without distortions. The rising front is only 5 ms long. GAEC detects double talk condition both in full band and in all relevant sub-bands.](image)
Figure 5.12. SND OUT signal with input double talk signal on –35 dBm (73 in units) passes, but with visible ±2 dB distortions. The rising front is 15 ms long. GAEC detects double talk condition both in full band and in all relevant sub-bands.

Figure 5.13. SND OUT signal with input double talk signal on –40 dBm (41 in units) passes, but the signal is attenuated about 6 dB, and there are visible ±2 dB distortions. The rising front is 25 ms long. GAEC detects double talk only in sub-bands responsible for 800 and 4000 Hz harmonics.

Figure 5.14. SND OUT signal contains only principal 800 Hz harmonic of the input double talk signal on –45 dBm (23 in units). The signal is attenuated about 6 dB, and there are strong visible ±4 dB distortions. The rising front is 25 ms longer. GAEC does not detect double talk condition (the actual threshold value for this case is –42 dBm).
Figure 5.15. SND OUT signal contains same traces of the input double talk signal on –50 dBm (13 in units). The signal is heavy attenuated, and although it is somewhat audible by a human ear, its quality is marginal. GAEC does not detect double talk condition on this input signal.

The signal used in this test is representative for simulation of a vowel as double talk signal.
6 Characterization of performance on voice signals

Depending on the specifics of the RCV voice excitation signal, performance of any AEC differs significantly. Language, age, gender, emotional context, pitch variability, loudness, background noise level and spectrum are between the major parameters affecting an AEC convergence. For example, softer voices are harder to converge on, voices with stronger pitch variation are easier, languages with many strong unvoiced phonemes (as Hebrew/Arabic) are easier, etc. It is not possible to cover all possible cases in this datasheet document.

A sentence in English of approximately 8 seconds, pronounced by a male speaker in his mid 20s, with –20…-10 dBm loudness, and with relatively low pitch variations is taken as an example of typical single talk speech segment. To evaluate GAEC performance on longer intervals of time, the same sentence is repeated over and over again.

A bit-exact simulation of a typical GAEC performance in a room with $T_{60} = 520$ms is presented. The echo path has the internal to speakerphone resonance modes (first 15ms) attenuated by 10 dB (relative to the real measured echo path), what would correspond to a relatively decent acoustic design. The loudspeaker and other electro-acoustic circuitry are assumed to be linear (less than 1% distortions) for the purpose of testing GAEC alone.

The simulation environment mixes pink noise on the –65 dBm level. The noise reduction circuitry is switched off because it behaves unrealistically well on artificial noises.

GAEC echo tail was chosen as 350 ms. The pCFg->sTCLst was configured as 40 dB.

6.1 Single talk convergence

![Graph showing the convergence of GAEC on a typical voice signal.](image)

Figure 6.1. Convergence of GAEC on a typical voice signal is stable and relatively fast.
Figure 6.2. The residual error on the repetitions of the input excitation signal. The TCLwst, as conservatively accessed by GAEC itself, was about 20 dB at the end of the first repetition, and 37 dB at the end of the last 5th repetition.

### 6.2 Double talk detection and processing

To test double talk processing, a known signal of a female speaker was added to SND signal as a double talk, between 3.1 and 4.8 seconds of each repetition.

Figure 6.3. The RCV and double talk (DT) signal energies. The Double talk signal is comprised of a breathing sound (3.1…3.8s), and several words, each starting with an unvoiced phoneme of low energy, followed by vowels with increasing energy (-45, -38, -25 dB). The final signal is a spike at 4.75s.
Figure 6.4. The spectrogram of the input double talk signal.

Figure 6.5. Spectrogram of the RCV signal during double talk period.

### 6.2.1 Double talk in early stage of convergence

TCL\textsubscript{wst}, estimated by GAEC itself, is 10.5 dB by the beginning of double talk. The processing of double talk is typical to a basic AEC: only low harmonics of the strongest signal are passed (even they are distorted), the rest is clipped off.

Figure 6.6. Double talk energies.
Figure 6.7. Spectrogram of the SND OUT signal. Only a trace of a spike at 3.5s, and only low frequency harmonics of the 4.6s and 4.75 signals.

Figure 6.8. The RCV OUT signal is passed with minor alterations. Most of those fall around 4.6 seconds, where low-frequency content is attenuated to give way to SND signal.

6.2.2 Double talk on the 2nd repetition
TCLwst, estimated by GAEC itself, is 24.5 dB by the beginning of double talk. All signals are passed, but the low-level double talk signals are still heavily attenuated and quite distorted. The spectrum of the breathing sound after the spike is realistic, but preceding inhaling sound is clipped. The high-frequency harmonics of all sounds start to become audible and visible on the spectrogram.
Figure 6.9. Energies on the second repetition.

Figure 6.10. Compare the SND OUT spectrogram on 2ns repetition with the spectrogram of the source double talk signal to see the results of post-processing.

Figure 6.11. RCV OUT signal undergoes attenuation in the same frequency-time areas where SND OUT signal is of maximal amplitude.
6.2.3 Further improvements.
Starting from the double talk on the 3rd repetition, the energy curve for the SND OUT follows double talk energy very closely. Even the starting inhaling sound is processed correctly. TCLwst, estimated by GAEC itself, is 31.5 dB by the beginning of double talk for the 3rd repetition, 34.5 – 4th, 37.5 – 5th.

The improvements are mostly in the increased precision of the spectrum transferring, and in the low-energy unvoiced phonemes.

The absolute value of TCLwst, achieved by GAEC, shall be related to the configured pCfg->sTCLst for making conclusions about double talk transparency. The difference between those parameters is the major factor.

![Figure 6.12. Energies on 3rd round.](image1)

![Figure 6.13. SND OUT spectrogram on the 3rd repetition.](image2)
Figure 6.14. SND OUT spectrogram on 4\textsuperscript{th} repetition.

Figure 6.15. SND OUT spectrogram on 5\textsuperscript{th} repetition.

Figure 6.16. RCV OUT spectrogram on 5\textsuperscript{th} repetition
6.3 Speaker Change

One of known deficiencies of an AEC is its dependency on the spectral properties of a given far-end talker. Usually, an AEC converges somehow while a far-end talker is talking, and it adapts to the pitch and spectral envelope of his/her voice. If somebody else starts talking on the far-end side (as it happens during conferences), an AEC needs to re-converge. During initial phase, an AEC may interpret the increased residual echo as a double talk signal.

To test GAEC, 3 repetitions of the same test sequence of male voice in English were used, followed by a sentence of a female person with different pitch, timbre and language.

If an AEC fails this test, there is extra signal on SND OUT, with clearly audible new pitch, which is visible on the signal’s spectrogram.

![Figure 6.17. Energies during speaker change. TCLwst (adaptive plus switched) always stays above pCfg->sTCLst (=40 dB). Double talk condition is not detected.](image)

![Figure 6.18. The spectrum of the RCV OUT signal shows that the spectrums before and after 25 seconds mark are very different, and pitch is more than twice higher.](image)
Figure 6.19. There is no trace of new pitch in the SND OUT signal after 25s, as seen on the spectrogram.

Tests passed.
7 Characterization of performance in a speakerphone

7.1 Setup #1

GAEC was tested as a part of a real speakerphone with a loudspeaker of marginal quality (<$1 price), μ-law codecs, and a simple (two parts only), strongly resonating housing. The speakerphone required a carefully designed set of equalizing filters to compensate its ±10 dB variations in frequency response in both RCV and SND direction.

All incoming and outgoing signals were recorded. The results of testing are consistent with off-line simulation, and they show the importance of having a quality loudspeaker in non-resonating housing.

7.1.1 Convergence on noise

For this speakerphone, the recommended echo tail length is between 150 and 250ms. Increase to 350…400 ms would lead to unjustified spending of MIPS and memory. This phone is not likely to be able to deliver good performance in rooms with $T_{60}$ longer than 400ms.

7.1.2 Convergence on voice

In this test, GAEC was instructed to keep RCV signal unmodified; no noise was added. The room with $T_{60} = 520$ms was used for testing. The SND IN noise level was about −50…−55 dBm and slightly fluctuating. The far-end person was using a handset and its noise level is lower, about −60 dBm.
Figure 7.2. GAEC with echo tail of 200ms converges within 2…3 seconds on a real voice signal to TCLwst of 20 dB and above. Yet, if the RCV signal is very strong and causes significant inter-modular distortions in the loudspeaker, the residual error (after echo cancellation) ERR jumps up. After noise reduction, the level of SND OUT signal is lowered to –60±5 dBm level (due to noise reduction circuitry), to be better stabilized later on.

It is evident that, after 4s time point, the level of residual echo (generally low) heavily depends on the RCV OUT signal. Whenever RCV OUT signal level approaches or exceeds –10 dBm level, the residual error tends to increase out of proportion. Zooming on 6.6s time point shows the necessary details of loudspeaker’s influence on GAEC performance.

The figures below show the correspondence of RCV OUT and ERR (residual error) spectrums.

Figure 7.3. The spectrum of RCV OUT signal between 6.55s and 6.65s has strong formants at about 500, 2000 and 2800 Hz.
Whenever the RCV OUT excitation signal has at least two strong components – one of low frequency, causing large voice-coil excursions (\(\sim 1/f^2\)), and a high frequency component, inter-modulation distortions follow. The residual echo sounds as a sudden crack, which has ‘strange’ spectrum envelope, but the same pitch as RCV OUT. Far-end users perceive this signal as a very annoying echo if it is passed through.

GAEC does not cancel such echoes, but predicts their power envelope and mask excessive echoes by inserting extra attenuation in SND direction. If a low-level double talk signal happens at the same time, it becomes completely suppressed, and even high-level double talk becomes audibly distorted.

Theoretically, mild soft non-linearity of quality loudspeakers can be identified, predicted and corrected with a mirror filter. Practically, that may be not true in the case of cheap loudspeakers, whose behavior tends to be far less repeatable as soon as they move into the area of severe distortions.

7.1.3 Subjective testing

Speakerphones with GAEC (90ms echo tail) were made available for subjective testing to a customer. The customer’s employees were asked to use it as frequently as possibly on both local and long distance call, in both hands-free mode, and as a conferencing phone. During each call, whenever appropriate, users asked the far-end person to express his or her opinion on the voice quality. It is recognized that these users do not represent a group of trained experts in assessing MOS voice quality.

- No complains on the speakerphone’s voice quality, of whatsoever nature, have been raised during the field test period.
- One of long-distance far-end parties clearly preferred the GAEC-equipped speakerphone to a popular, much more expensive, conferencing phone, which was routinely used before in the same conditions.
- Many far-end parties could not tell hands-free mode from handset mode.
- All near-end users appreciated speakerphone’s sound.

The overall test response gave an indication that GAEC’s quality exceeded the users’ expectation.

7.2 Setup #2

A reference setup using ‘C5510 DSK, external loudspeaker and external electret microphone was used. The components were chosen to be of low- to mid grade quality. No professional level components (which can not be reproduced in a commercial speakerphone design) were used. The loudspeakers were compensated to have ±3dB flat frequency response in 250Hz … 3400 Hz range.
Figure 7.5. The acoustic components of setup.

It was found that

- The quality of a mid-level bookshelf loudspeaker (estimated cost of $15) exceeds the non-linearity requirement of speakerphones.
- Even the quality of a cheap entry-level PC 3W loudspeaker is sufficient to obtain full-duplex sound. The quality of those loudspeakers is higher than of used in a typical speakerphone.
- The better is the loudspeaker, the more natural and easily intelligible is the sound of the far-end party.

7.2.1 Convergence on noise

![Graph showing the convergence limits for the entry-level PC loudspeakers with sufficient echo tail length.](image)

Figure 7.6. The convergence limits for the entry-level PC loudspeakers with sufficient echo tail length. Observing this picture, the pCfg->sRcvDistThr shall be set to –13 dBm.
7.2.2 Convergence on voice

Single talk, male speaker RCV signal of average level was used for testing. The call was incoming. A bookshelf loudspeaker was used.

\[ \text{Figure 7.7. Initial phase of convergence on voice. GAEC is capable of pre-adapting on preceding noise signal, however, correct assessment of the noise level is delayed.} \]

\[ \text{Figure 7.8. The first 25 seconds of the call (near end speaker was instructed to keep silence and stay still). The residual echo becomes almost suppressed down to noise floor by adaptive means after 10…15 seconds.} \]

7.2.3 Subjective testing.

The results of subjective testing showed that achieving of perceptually transparent sound with GAEC is possible.

- Most of the far-end parties could not recognize that the near-end party is in handsfree mode.
- Only musicians or people with keen hearing reported that they recognized handsfree mode by noticing the reverberating effect on the signal of the near-end party (not the echo of their voice). However, the same people do not find it objectionable, and moreover, tend to prefer flat microphone response as sounding more natural, opposing to typical +3dB per octave pre-emphasized response, preferred by non-musicians, which makes reverberating effect less audible.
• The near-end reverberation effect is less audible if the mouth-microphone distance is within ITU-T G.167 nominal 0.5m distance.
• No complains on the speakerphone’s voice quality, of whatsoever nature, have been raised during the field test period.

7.3 Conclusions

GAEC is capable of removing acoustic echoes and providing output signal with minimal clipping for both far- and near-end users even if inexpensive electro-acoustic components are used throughout a speakerphone, but their quality sets hard limits to the overall quality of speakerphone.

It was mentioned that the loudspeaker’s distortions have major impact on the overall performance of entire speakerphone. The amount of adaptive terminal coupling loss (TCLwst), which can be achieved with a particular design, limits the ability of GAEC to recognize and correctly process low-level double talk signal, thus decreasing double-talk range. If the maximum achievable adaptive TCLwst and the target configured pCfg->TCLst differ significantly (20+ dB), GAEC may suppress double talk signal, and insert high attenuation in both direction, with obvious impact on the transparency of a speakerphone and its overall quality. It might be impossible to achieve the theoretical limit 50 dB of TCLwst in a real speakerphone in a real room with reverberation time about 400 ms, but the closer this limit is approached, the better will be the resulting sound quality for both near-and far-end parties.

To achieve a good hands-free performance with a phone employing GAEC, its designers are advised to pay extra attention on the acoustic design and use components of balanced quality.
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.