LIMITS ON ECHO RETURN LOSS ENHANCEMENT ON A VOICE CODED SPEECH SIGNAL

Mark Rages and K.C. Ho

Department of Electrical and Computer Engineering
University of Missouri - Columbia
Columbia, MO 65211
markrages@mlug.missouri.edu

ABSTRACT

Acoustic echo cancellation is a desirable feature for small cellular handsets that have significant acoustic coupling from speaker to microphone. Current echo cancellation techniques prefer an echo canceller to be placed at the wireless handset. Because echo cancellation requires much processing power it is beneficial to move the echo canceller to the base station, where power consumption will not be a constraint. This paper examines the feasibility of performing linear echo cancellation on a signal coming from a mobile at the base station. The mobile signal at the base station contains speech coding which may degrade the effectiveness of echo cancellation. The performance of linear echo cancellation on mobile echo encoded and decoded by a low bit rate speech coders operating at 4 kbits/sec to 13 kBits/sec, including GSM full rate, G.723a, G.729, and AMR speech coders, was investigated. The best echo return loss enhancement achieved depends on the bitrate of the coder, and is usually less than 10 dB for the low bit rate operating range examined.

1. INTRODUCTION

The advent of small wireless handsets with increased acoustic coupling between the speaker and microphone has made acoustic echoes generated within the handset a problem for high-quality speech communication. Echo cancellation can be used within the handset to improve voice quality. As shown in Fig. 1, the adaptive filter $H(z)$ models the acoustic echo path h° and produces an echo estimate which is subtracted from the microphone signal before it is encoded and transmitted to the other user. Performing echo cancellation in the handset is expensive in terms of power consumption and processor cost.

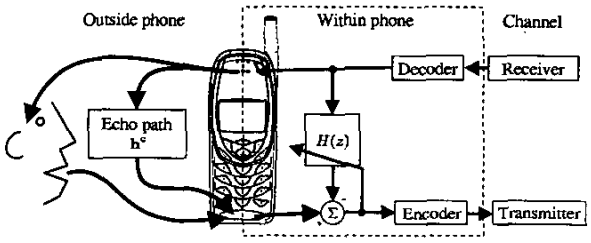

Figure 1: Echo cancellation within the handset

One way to alleviate the extra processing required by the handset is to allocate the processing of echo cancellation to the base station, where signal processing resources are not as expensive.

One difference in preprocessing echo cancellation in the base station instead of the handset is that the mobile signal received at the base station has passed through two speech codecs as shown in Figure 2. The codecs introduce nonlinear distortion to the mobile signal, and this may reduce the effectiveness of echo cancellation.

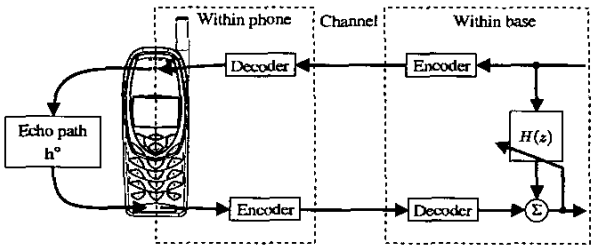

Figure 2: Echo cancellation in the base station

The nonlinearity from speech coding is mainly due to the excitation approximations and quantization. Tandem speech coding, where two speech codecs operate on a signal in cascade fashion, as in the case of mobile communication, can cause an even more severe effect upon echo cancellation.

This exploratory study will examine the limits on the performance of linear echo cancellation when one and two speech codecs are present in the echo path. The results will provide some understanding about the effect of speech coding on linear echo cancellation. Section 2 examines the distortions added to the signal by speech codecs. Section 3 discusses linear echo cancellation. Section 4 describes the tests used to measure the echo return loss enhancement. Section 5 presents some experimental results. Section 6 gives an interpretation of the results. Section 7 is the summary.

2. SPEECH CODING

The most popular speech coders are based on the speech production model. In this model, the speech is divided into two parts: the excitation (from the larynx) and the formant filter (from the throat and nasal passages). The excitation is usually modeled as a series

0-7803-7523-8/02/$17.00 ©2002 IEEE

of pulses, and the filter is modeled as a linear all-pole filter, both for simplicity and for perceptual reasons.

Because speech is time-varying, not stationary, speech coders divide the speech into small frames of 10 - 30 msec. The analysis and coding is then carried out per frame.

One of the most common speech coders is the code-excited linear prediction (CELP) coder. This coder first uses linear prediction to model the formant filter, which removes short-term correlation from the signal. The poles of this filter are often transformed into line spectral frequencies for transmission.

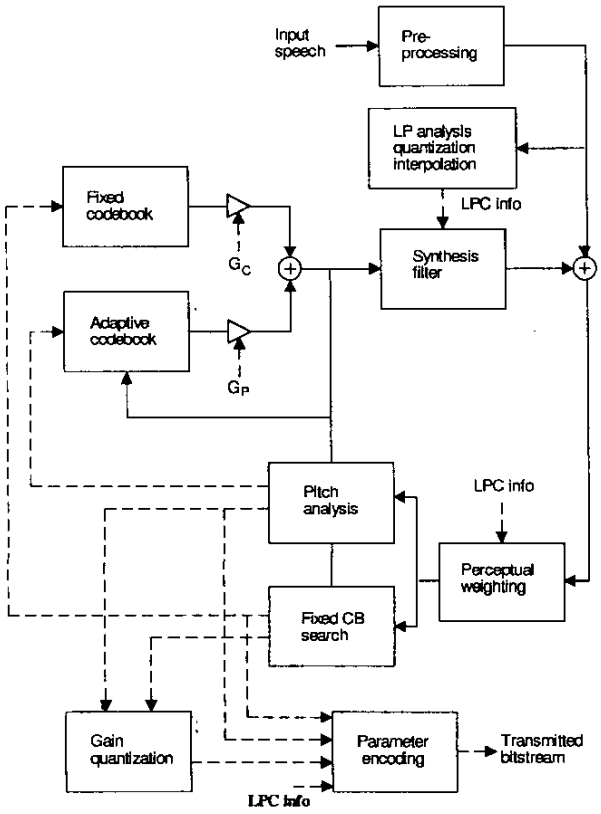

Figure 3: Block diagram of G.729 speech coder[1]

The excitation remaining after linear prediction is then coded. The CELP coder belongs to the class of analysis-by-synthesis coders, which find the best excitation by reconstructing within the encoder the same signal that the decoder will decode. The decoded response is calculated for each entry in the excitation codebook. The best codebook entry, usually the one that gives the lowest perceptually-weighted mean-square error between the original and decoded speech, is chosen and the codebook index is transmitted.

Another feature of most CELP systems that improves on simple codebook lookup is the adaptive excitation contribution. The pitch of the speech is found by searching the cross-correlation of the linear prediction residual. The pitch delay and magnitude are used to predict the excitation from the previous pitch cycle. The codebook value is added to this to generate the final excitation in the decoder.

The block diagram of a representative CELP codec, the ITU Standard G.729 is shown in Figure 3. This codec has a bit rate of 8.0 kbit/s.

3. LINEAR ECHO CANCELLATION

Echo cancellation in the base station will have two speech codecs in tandem as shown in Fig. 2.

When two codecs operate in tandem, in general the signal quality becomes worse. However, it may not be necessary to operate codecs in tandem for echo cancellation. Figure 4 shows a possible tandem configuration, and Figure 5 shows how the tandeming may be eliminated for echo cancellation by feeding the decoded speech signal to the filter $H(z)$. While the single codec case appears to have an additional decoder block, this function is usually provided as a side effect of the analysis-by-synthesis speech coder. In any case, decoding the signal is not computationally expensive.

The simplest method of echo cancellation is the least mean square (LMS) algorithm.

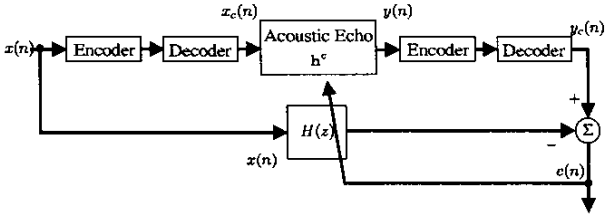

Figure 4: Echo cancellation with a tandem of codecs in the echo path

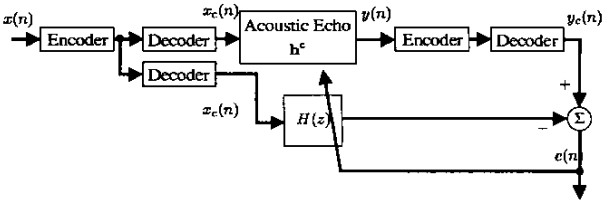

Figure 5: Echo cancellation with a single codec in the echo path

Let the incoming speech samples be $x(n)$ and the signal from the microphone be $y(n)$. The signal to be transmitted to the far end will be denoted as $e(n)$. This quantity is called an error signal, a term borrowed from adaptive filter theory.

We are interested in cancelling the echo in the mobile signal. Given a chosen filter length N for $H(z)$, we define the signal vector $\mathbf{x}(n) = [x(n), x(n-1), ..., x(n-N+1)]^T$.

The signal $x(n)$ is first encoded in the base station for transmission to the mobile. The speech data is decoded in the wireless handset and becomes the acoustic energy $x_c(n)$ at the handset loudspeaker. Due to insufficient acoustic decoupling, some of this energy leaks through the handset and enters the microphone. If the user of the handset is not talking, then $y(n)$ consists only of the echo. For purposes of simplicity, the echo path is modelled as a linear filter $h^c(n)$. (Note that the actual handset acoustic response

is not linear! [2][3]) Then $y(n) = \sum_{i=0}^{N-1} x_c(n)h^\circ(n-i)$. $y(n)$ is encoded in the mobile and then decoded at the base station to form $y_c(n)$.

The problem of linear echo cancellation at the base station is to find an equivalent impulse response estimate, denoted as $\mathbf{h}^\circ = [h^\circ(0), h^\circ(1), ..., h^\circ(N-1)]$, of the combined speech codecs and echo path $\mathbf{h}^\circ(\mathbf{n})$, given only the transmitted signal $x(n)$ and recovered signal $y_c(n)$.

For every estimate of the equivalent echo path $\mathbf{h}$ we can write the error signal as

$$e(n) = y_c(n) - \mathbf{h}^T\mathbf{x}(n) \quad (1)$$

The LMS algorithm updates $\mathbf{h}$ by

$$\mathbf{h}(n+1) = \mathbf{h}(n) + 2\mu e(n)\mathbf{x}(n) \quad (2)$$

where μ is the step size and $\mathbf{h}(n)$ is the echo path impulse response estimate at time n.

In this study, we examine the best achievable results. It is well known that the LMS algorithm converges to the Wiener solution. For a long enough data record of length L, the Wiener solution is given by

$$\mathbf{h} = \mathbf{R}^{-1}\mathbf{p} \quad (3)$$

where

$$\mathbf{R} = \sum_{n=0}^{L-1} \mathbf{x}(n)\mathbf{x}(n)^T \quad (4)$$

and

$$\mathbf{p} = \sum_{n=0}^{L-1} y_c(n)\mathbf{x}(n) \quad (5)$$

A measure of the effectiveness of an echo cancellation system is the echo return loss, calculated by taking the ratio of echo power before and after echo cancellation:

$$ERLE(dB) = -10 \cdot \log_{10} \frac{\sum_{n=0}^{L-1} e(n)^2}{\sum_{n=0}^{L-1} y_c(n)^2} \quad (6)$$

The residual echo $e(n)$ has two components, a linear and a nonlinear part. Linear echo cancellation can only remove the linear part of the echo and the noinlinear part remains in e(n).

4. ENHANCEMENT MEASUREMENT METHOD

Experiments were performed to examine the ERLE when the echo path contains speech codecs. Simulations used thirty seconds of male speech and thirty seconds of female speech data taken from the TIMIT[4] speech corpus.

For each combination of codec, the echo return loss enhancement was measured using (6). To establish a limit on the amount of return loss, the simplest possible echo path $\mathbf{h}^\circ$ was used, an impulse of magnitude 0.5 at location $N/2$, where $N = 100$.

To measure the echo return loss enhancement with voice coding in the signal path, several steps were applied.

First, a dummy codec was used that simply copies signals from its input to its output. This was to verify that the echo canceler was working properly and to establish a reference of the maximum achievable ERLE.

Codecs were then inserted into the path, first singly, then in tandem. After the signal passed through a codec, it was normalized to contain the same energy at the output of the codec that it possessed at the input of the codec.

The codecs used for testing included the GSM full-rate codec[5], the adaptive multi-rate (AMR) codec[6], the G.723a codec[7], the G.729 codec[1], and the IS-641 codec[8]. These codecs were chosen for their popularity. The GSM codec is designed for mobile phone use at 13.2 kbit/s. The AMR codec is also intended for mobile phone use at bit rated between 4.75 and 12 kbit/s. G.723a is a 5.3 or 6.3 kbit/s codec designed for VoIP use. G.729 is an 8.0 kbit/s codec designed for multimedia applications. The AMR and G.723a codecs also have the interesting property of being multirate codecs. They allow testing the increasing limitations on echo cancellation with declining bitrates. These are all CELP codecs, with the exception of GSM. GSM is an older design known as regular pulse excitation, long term prediction (RPE-LTP).

5. RESULTS

Single codecs were tested at each bit rate. Table 1 and Table 2 show the results.

	Dummy 128 kbit/s	GSM 13.2 kbit/s	G723a 6.3 kbit/s	G723a 5.3 kbit/s	G729 8.0 kbit/s	IS641 7.4 kbit/s
Male	56.86	6.23	5.88	4.33	5.49	6.88
Female	57.91	7.48	8.87	7.52	7.78	8.83

Table 1: Echo return loss enhancement, dB ($N = 100$)

	AMR 4.75 kbit/s	AMR 5.15 kbit/s	AMR 5.9 kbit/s	AMR 6.7 kbit/s
Male	3.27	3.26	6.11	5.59
Female	7.07	6.69	7.56	8.81
	AMR 7.4 kbit/s	AMR 7.95 kbit/s	AMR 10.2 kbit/s	AMR 12.2 kbit/s
Male	7.35	5.93	10.30	11.07
Female	9.46	9.02	12.48	13.23

Table 2: Echo return loss enhancement, dB ($N = 100$)

Two codecs were tested in tandem at each of the varied bit rates. Both codecs in the tandem pair were tested at the same bit rate. Table 3 and Table 4 show the results.

	GSM 13.2 kbit/s	G723a 6.3 kbit/s	G723a 5.3 kbit/s	G729 8.0 kbit/s	IS-641 7.4 kbit/s
Male	4.16	2.96	2.16	2.61	2.89
Female	5.08	5.89	3.34	4.82	5.47

Table 3: Echo return loss enhancement of tandem codecs, dB ($N = 100$)

	AMR 4.75 kbit/s	AMR 5.15 kbit/s	AMR 5.9 kbit/s	AMR 6.7 kbit/s
Male	-1.34	-0.98	2.33	2.22
Female	4.34	3.54	4.53	4.46
	AMR 7.4 kbit/s	AMR 7.95 kbit/s	AMR 10.2 kbit/s	AMR 12.2 kbit/s
Male	3.88	3.74	6.46	7.40
Female	4.65	5.52	8.78	9.80

Table 4: Echo return loss enhancement of tandem codecs, dB ($N = 100$)

6. DISCUSSION OF RESULTS

For the single speech codec case, a codec causes the linear echo cancellation to decrease by about 50 dB, as shown in Table 1. This is probably due to the coding and quantization of the excitation parameters. The linear adaptive filter can only remove the linear component in the echo signal, and not the nonlinear part.

The tandem codecs are even worse than the single codecs. Some of the low bit-rate tandem codecs showed more echo power with echo cancellation than without it, which causes the ERLE measurement to be negative. This is possible because the linear echo cancellation can only cancel linear echoes perfectly. It fails when the echo path is nonlinear and time varying, such as is the case when a speech coder is inserted in the signal path. When the bit rate increased, the ERLE improved a little.

The study shows a need for future research. Possible topics include:

- Non-linear echo cancellation. Recent studies[2][3] have shown the acoustic response of the wireless handset is highly nonlinear, necessitating a nonlinear echo canceler. Such an echo canceler is more costly computationally than a linear canceler, putting even more pressure to move the canceler out of the handset and into the base station. Additionally, a non-linear echo canceler may be better able to deal with the speech coding artifacts.
- Instrumented codec. A modular CELP codec could be built that would allow measurement of the effects of different quantization upon echo cancellation. Perhaps an alternative analysis-by-synthesis criterion would allow better echo cancellation.

7. SUMMARY

The performance limitations of linear echo cancellation of an echo which has been coded by one or two speech coders has been examined and measured. The results indicate that linear echo cancellation is not enough to provide sufficient reduction in the echo level to improve voice quality, if the transmission link involves speech coding. Codecs with lower bit rates and tandem codecs will cause the most severe performance problems to a linear echo canceler. Additional research is needed to characterize the effects of speech coding on non-linear echo cancellation.

8. REFERENCES

[1] "Coding of speech at 8 kbit/s using conjugate-structure algebraic-excited linear-prediction," Standard ITU-T G.729, International Telecommunication Union, 1996.

[2] A. Fermo, A. Carini, and G. Sicuranza, "Analysis of different low complexity nonlinear filters for acoustic echo cancellation," in *Proceedings of the First International Workshop on Image and Signal Processing and Analysis*, pp. 261–266, 2000.

[3] J.-P. Costa, T. Pitarque, and E. Thierry, "Using orthogonal least squares identification for adaptive nonlinear filtering of GSM signals," in *ICASSP 97*, vol. 3, pp. 2397–2400, 1997.

[4] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," Standard LDC93S1, Linguistic Data Consortium, 1993.

[5] "Full rate speech transcoding," Standard GSM6.10, European Telecommunications Standards Institute, 1992.

[6] "Adaptive multi-rate (AMR) speech transcoding," Standard ETSI EN 301 704 V7.2.1 (2000-04), European Telecommunications Standards Institute, 2000.

[7] "Dual rate speech coder for multimedia communications," Standard ITU-T G.723.1, International Telecommunication Union, 1996.

[8] "TDMA radio interface, enhanced full-rate speech codec," Standard PN-3467, TIA/EIA, 1996.

[9] R. M. Storn, "Echo cancellation techniques for multimedia applications - A survey," Tech. Rep. TR-96-046, International Computer Science Institute, Berkeley, CA, 1996.

[10] "Digital network echo cancellers," Standard ITU-T G.168, International Telecommunication Union, 1996.