UTEC-CSc-77-090
Semi-Annual Technical Report

(12)

NOISE SUPPRESSION METHODS FOR ROBUST SPEECH PROCESSING

Contractor:	University of Utah
Effective Date:	1 October 1976
Expiration Date:	30 September 1978
Reporting Period:	1 October 1976 – 31 March 1977

Principal Investigator:	Dr. Steven F. Boll
Telephone:	(801) 581-8224

Sponsored by
Defense Advanced Research Projects Agency (DoD)
ARPA Order No. 3301
Monitored by Naval Research Laboratory
Under Contract No. N00173-77-C-0041

April 1977

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: UTEC-CSc-77-090	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): Noise Suppression Methods for Robust Speech Processing	5. TYPE OF REPORT & PERIOD COVERED: Semi-Annual Technical rpt. 1 Oct 1976 – 31 March 1977
6. PERFORMING ORG. REPORT NUMBER	
7. AUTHOR(s): Dr. Steven F. Boll	8. CONTRACT OR GRANT NUMBER(s): N00173-77-C-0041 ARPA Order 3301
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of Utah, Computer Science Department, Salt Lake City, Utah 84112	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Project: 76-RPA-3301
11. CONTROLLING OFFICE NAME AND ADDRESS: Defense Advanced Research Project Agency (DoD), 1400 Wilson Blvd., Washington, D. C. 22209	12. REPORT DATE: April 1977 13. NUMBER OF PAGES: 77
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): Naval Research Laboratory, 4555 Overlook Avenue, S. W., Mail Code 2415-A.M. (12) 84p.	15. SECURITY CLASS. (of this report): Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

Same

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Digital noise suppression; Linear Predictive Coding; Narrow band coded speech; Adaptive noise cancellation; Wiener filtering; Power spectrum; Autocorrelation.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Robust speech processing in practical operating environments requires effective environmental and processor noise suppression. This report describes the technical findings and accomplishments during this reporting period for the research program funded to develop real time, compressed speech analysis-synthesis algorithms whose performance is invariant under signal contamination. Fulfillment of this requirement is necessary to insure reliable secure compressed speech transmission within realistic

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102 LF 014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

404 749

20. ABSTRACT con't.

military command and control environments. Overall contributions resulting from this research program include the understanding of how environmental noise degrades narrow band, coded speech, development of appropriate real time noise suppression algorithms, and development of speech parameter identification methods that consider signal contamination as a fundamental element in the estimation process. Through the appropriate integration of developed noise suppression-parameter identification algorithms, specifications for robust speech processing algorithms will be provided.

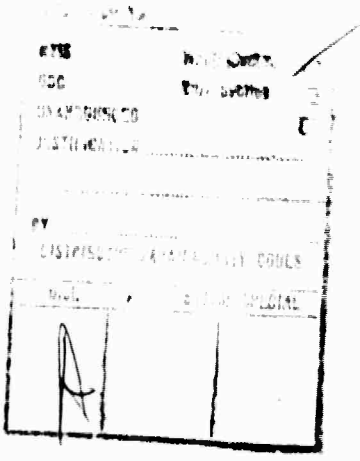


TABLE OF CONTENTS

SECTION I

Summary of Program for Reporting Period

A. Introduction

1. This section summarizes the objectives, tasks and results of the research program for the period 1 Oct. 1976 through 31 March 1977. Detailed descriptions are provided in the remaining sections.

B. Objectives

1. Accumulate, digitize and categorize a representative data base consisting of clean speech, noise and noisy speech needed for measuring speech analysis algorithm performance and noise suppression algorithm effectiveness.
2. Investigate the time and frequency relationships between speech and additive noise as well as the corresponding analysis parameter variations resulting from the analysis of the noisy speech.
3. Develop integrated noise suppression speech analysis algorithms which improve the quality and intelligibility of coded speech by modifying the analysis equations to explicitly account for and thereby suppress the noise.
4. Develop preprocessing noise suppression algorithms using two microphone inputs which will improve the

signal-to-noise ratio prior to vocoder input.

C. Tasks Undertaken and Results

1. A data base was digitized from recordings of laboratory noise; speech and noise spoken in a quiet room, office and helicopter environments; and audio test sentences used in NSA consortium testing. In addition, utility programs for measuring and scaling data were developed.

2. Interactive display and playback programs were developed for comparing speech spectra, correlations, and analysis parameters for various signal-to-noise ratios and environments.

   a. Our research determined that the spectral and temporal distortions resulting from LPC analysis of noisy speech include:

      (1) Widened formant bandwidths.

      (2) Shifted formant center frequencies.

      (3) Low energy formants partially or completely obscured by noise floor.

      (4) Overall decrease in spectral dynamic range.

      (5) Increase in peak factor of voiced synthetic speech with the accompanied increase in annoying "buzzy" quality.

   b. In addition the research determined that the short time crosscorrelations between speech

and broadband Gaussian noise do not average to zero. Thus it is incorrect to assume that speech and Gaussian white noise are uncorrelated during short time analysis intervals.

c. Using the LPC parameter comparison program it was demonstrated that as the signal-to-noise ratio decreases the spectral distortion as measured by the Gray and Markel distance measure increases as follows:

SNR(dB)	Cosh Distance(dB)	Cepstral Distance(dB)
40	0	0
30	2.5	2.3
20	5.4	4.5
10	7.4	6.2
0	9.2	7.3
-10	10.2	8.1

3. An expanded speech analysis method was developed called "Predictive Noise Cancellation" to suppress noise by modifying the speech autocorrelations prior to LPC coefficient calculation. Estimates of current noise values were adaptively predicted from long term noise statistics taken during non-speech intervals.

   a. Predictive Noise Cancellation offers the

advantages of:

(1) Uses procedures which are currently available in real time LPC vocoders.

(2) The method results in a stable synthesis filter.

(3) Background noise is reduced by 10 to 20 dB.

However it has two major disadvantages:

(4) The method is dependent upon the phase of the signals processed.

(5) The estimate of the noise-signal correlation filter is corrupted when speech is present.

4. A two microphone input noise cancellation algorithm which has been used effectively in the areas of antenna side-lobe attenuation and data channel equilization was implemented and calibrated to determine its effectiveness in reducing noise prior to vocoder input. From one microphone is recorded speech plus noise and from the other, a correlated noise signal.

   a. Preliminary results demonstrated that the method will remove broadband noise which has been digitally added to speech, by subtracting the second adaptively filtered noise channel from the noisy speech.

b. Signal-to-noise improvements up to 40dB were measured.

E. Future Efforts

1. Based upon the success of the dual input adaptive noise cancellation algorithm for removing digitally added laboratory noise, the method will be applied to the removal of noise found in office and helicopter environments.

2. The inadequacies of the Predictive Noise Cancellation method can be removed by using a frequency domain spectral averaging technique. Although this technique requires Fourier transforms, it appears to be implementable in real time, applicable to other vocoder forms such as channel or homomorphic, and have better noise cancellation properties.

3. It can be shown that an all-pole process corrupted by additive Gaussian noise can be modeled as a pole-zero process. An investigation is now under way to determine whether algorithms for estimating pole-zero processes can be adapted to find the predictor coefficients corresponding to the underlying clean speech.

III. RESEARCH ACTIVITIES

SECTION I

Summary of Overall Research Program

Program Objectives

Primary Objective

To develop robust speech processes, based upon the integration of digital noise suppression methods and narrow band speech analysis-synthesis methods, capable of realizing practical, real time methods for effectively processing speech recorded in practical operating environments.

Support Objectives

To specify noise suppression methods for robust speech processing will require the following tasks:

1. Accumulation and categorization of signal contamination associated with practical operating environments.
2. Categorization of currently used speech processing algorithm performance, e.g. Linear Predictive Coding, (LPC) in these operating environments.
3. Development of real time noise suppression algorithms and categorization of their effectiveness in reducing signal contamination.
4. Development of new or modified speech analysis

algorithms which can effectively extract acoustic parameters from contaminated speech.

5. Specification for robust algorithms through the integration of noise suppression-parameter identification algorithms.
6. Documentation and demonstration of robust algorithm performance using contaminated speech.

Research Plan

The research program consists of three parallel but interactive subprograms. These programs are described as (1) Operating Environment Understanding; (2) Noise Suppression Algorithm Development; (3) Speech Processing Algorithm Development. The study is applied to contaminated signals generated both in the laboratory as well as actual operating environment. In addition, examples of the contaminated signals have been provided for the program by the National Security Agency (NSA). Using this data base, the program addresses signal contaminations associated with realistic military environments.

Research Approach

The program is broken down into four phases. Within each phase the parallel tasks of environment understanding, noise suppression algorithm development, and speech

processing algorithm development are carried out.

Initially the characterization of the environments and how they effect the speech analysis methods must be understood. This characterization is done in Phase I of the program. Next, it must be determined how the current ARPA-NSC speech processing algorithms perform in the environments. In Phase II the quantitative performance of the algorithms will be measured using both contaminated and uncontaminated speech. These measures are obtained by comparing the output analysis acoustic parameters (such as pitch, voicing, gain, etc.) estimated using both contaminated and non-contaminated signals as well as spectral deviations. Examples of these comparisons are presented in this report.

After having characterized the signal contamination, as well as the algorithm's response to the contamination, decisions will be made as to how to effectively and efficiently suppress or eliminate the noise using algorithms either already implemented or currently being developed. Thus in parallel with the above tasks, will be the development of noise suppression algorithms.

In Phase III the choice of which algorithm to use based upon which type of contamination is present will be made. After the appropriate integration of noise suppression and speech parameter identification algorithms, the resulting

system's performance to undistorted and distorted speech will be categorized and demonstrated.

In Phase IV the implementation requirements needed to interface the resulting robust algorithms to the ARPA network speech communication system will be determined and specified. Below is a summary of the task orderings.

Summary of Tasks

Phase I Noise Characterization and Processor Implementation

1. Accumulate and categorize signal contamination data base.
2. Initiate theoretical investigation of parameter estimation techniques based on degraded input speech.
3. Develop appropriate utility programs needed to manipulate data and display essential features.

Phase II Measurement of Algorithm Performance Using Contaminated Speech

1. Determine the performance of present, unmodified speech compression algorithms using contaminated speech and categorize results.
2. Determine the performance of noise suppression algorithms to undistorted and distorted speech and

categorize results.

3. Continue theoretical investigations of parameter estimation techniques based upon degraded speech for suppressing the known noise environments categorized in Phase I in order to compensate for present vocoder limitations.

Phase III Specifications for Robust Speech Processing Algorithms

1. For each operating environment, determine the appropriate integration of noise suppression-parameter identification algorithm.
2. Categorize the resulting robust system's performance to undistorted and distorted speech.
3. Demonstrate and document robust system improvement and performance for the different operating environments.

Phase IV System Implementation and Protocol Specifications

1. Determine and specify implementation requirements needed to interface robust algorithms to ARPA network speech communication system.

SECTION II

Generation and Calibration of Data Base

A. Objectives

1. Accumulate, digitize and categorize a representative data base for measuring speech analysis algorithm performance and noise suppression algorithm effectiveness.

B. Approach

1. Digitize laboratory generated noise files which model components found in actual operating environments.
    a. Wide band uncorrelated Gaussian noise
    b. Wide band correlated periodic noise
2. Digitize speech and noise recorded in both quiet, ideal and noisy actual operating environments.
    a. "Clean" speech having negligible additive noise component.
    b. Speech recorded live in helicopter cockpit.
    c. Speech recorded live in normal office environment.
3. Obtain and digitize speech used by National Security Agency as part of consortium test of narrow band devices.

4. Calibration of laboratory noise.

   a. Measure average energy of noise and clean speech files.

   b. Specify desired signal-to-noise ratio, SNR.

   c. Scale noise files by appropriate gain and add to clean speech files.

   d. Generate contaminated speech files having SNR ranging from -10dB to 40dB.

5. Specify and categorize type of signal contamination

   a. Laboratory noise digitally added to clean speech.

   b. Field noise digitally added to clean speech.

   c. Field noise acoustically added to speech (true field conditions).

C. Tasks

1. Generation of laboratory noise.

   a. The output from an analog noise generation was digitized and recorded at sampling rates of 6.67KHZ, 8.0KHZ and 10.0KHZ.

   b. The output of a square wave generator was digitized and recorded at rates of 6.67KHZ, 8.0KHZ and 10.0KHZ.

      (1) fundamental frequency was both fixed at 400 HZ and varied linearly from 400 HZ to approximately 1000 HZ.

2. Generation of field data.

a. Using a portable stereo cassette recorder and a Sony directional microphone live stereo recordings were made in following environments:

(1) Sensory Information Processing-group "Quiet Room"

a. Ambient noise level = 27dB

b. Speech recorded in this environment was used as noise-free "clean" text.

b. Computer Science Department Office

(1) Ambient noise level = 65dB

(2) This speech data represented the office environment.

c. Ramjet Helicopter Cockpit

(1) Ambient noise level = 105dB.

(2) This speech data represented the helicopter environment.

3. Recordings of National Security Agency's consortium audio test tapes.

a. Three speakers in three environments were recorded.

(1) Environments: Quiet, Office, Helicopter.

b. Helicopter noise without speech was also recorded.

c. Data was filtered at 3.2KHZ and sampled at 6.67KHZ.

4. Development of utility programs needed to measure signal energy and adjust signal-to-noise ratios.

a. Program for measuring signal energy:

(1) Program name: BWEGHT

(2) Program authors: W. Done, D. Pulsipher and J. Youngberg

(3) Program description.

BWEGHT

In analyzing the effects of noise on the various systems being tested, a measure is needed for the signal-to-noise ratio (SNR) that will quantify the degradation of various noise levels. The measure should also match the degradation the listener intuitively believes will occur at a given noise level. Because the final monitor of the systems being tested is the human ear, the measure selected is based on the B-weighting curve used for calibration of audio equipment. The B-curve is a member of a family of curves which, for specific ranges of energy, indicate sound energy levels throughout the auditory range which will be perceived as constant loudness levels. The approximation used for the B-curve is given by:

$$B(f) = \frac{7160.^2 f^2}{f^4 + 4.90256 \times 10^7 f^2 + 1.25440 \times 10^{12}} .$$

A plot of B(f), which represents a power spectra weighting function, is shown in Figure II.2. A program called BWEGHT has been written which performs the following:

1) Inputs a frame of speech (or noise) of width NW;

2) Windows that frame with a Hamming window (optional);

3) Calculate the DFT (of order NU) of the frame;

4) Finds the magnitude squared, S(f);

5) Multiply S(f) by the weighting function, B(f);

6) Calculate the energy in that frame of data, by

$$E_f = \sum_{j=0}^{\frac{N}{2}+1} P(f_j), \qquad f_j = \frac{(j-1)f_s}{N}$$

$$N = 2^{NU}$$

f_s = sampling frequency

$P(f) = S(f)B(f)$,

7) Proceed to the next frame of speech (or noise).

8) Calculate the average energy per frame by averaging the E_f found for each frame.

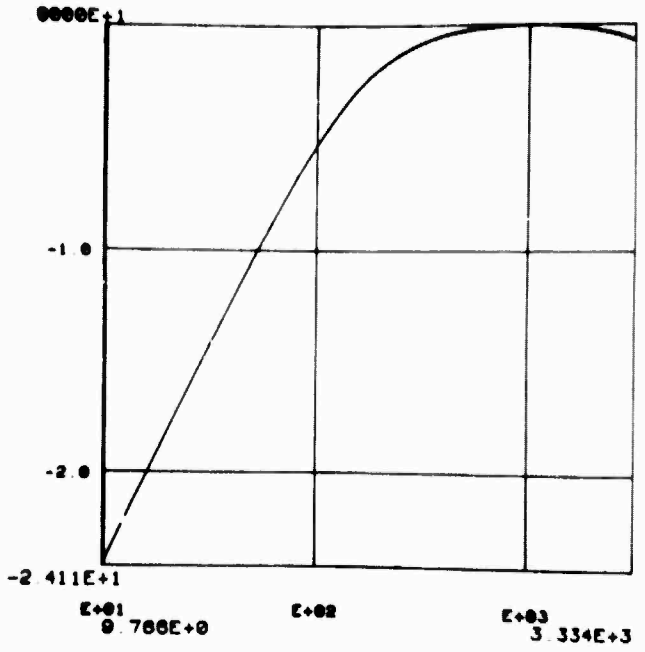
BWEGHT, then furnishes an average energy/frame for that passage. Typical parameter values are:

NW = 2048 (frame size)

NU = 13 (order of DFT)

(a)

log-log scale



(b)

Linear Scale

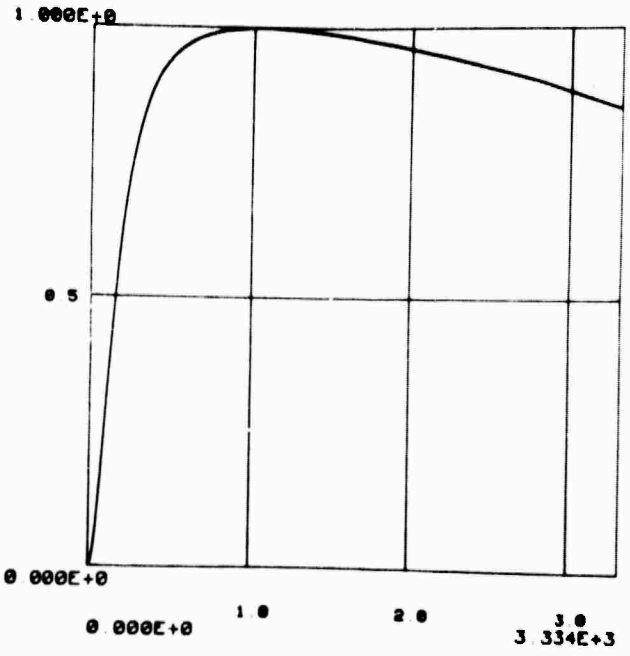


Figure II.2

B-Weighting Spectral Energy Curve

b. Program for adding known amounts of speech and noise:

(1) Program name: SPLUSN

(2) Program authors: W.Done, D. Pulsipher.

(3) Program description:

SPLUSN

Testing of noise suppression systems requires the use of data contaminated by known amounts of noise. For this reason, a program was designed which would scale a given noise file to achieve a certain signal-to-noise ratio when added to a speech file. This program, SPLUSN, produces an output file, x(k), according to either of the following equations (depending on whether the scaled noise or signal plus scaled noise is desired):

$$x(k) = c \cdot n(k)$$

$$x(k) = s(k) + c \cdot n(k)$$

where c is a constant.

The constant c can be entered in one step as a single number, or as function of E_s and E_n, the energy contained in the sequences s(k) and n(k), respectively, and S, the desired signal-to-noise ratio. The constant c is related to

these parameters by:

$$c = \left| \frac{E_s}{S \cdot E_n} \right|^{1/2}$$

where

$$S = 10^{S_d/10}$$

and S_d is the signal-to-noise ratio in dB. E_s and E_n are obtained from s(k) and n(k), respectively, by using the program BWEGHT described previously.

D. Results

1. A data base has now been recorded and digitized containing three types of data.

   a. Laboratory noise digitally added to speech.

      (1) Speech plus wide-band Gaussian noise with SNR under program control.

      (2) Speech plus periodic noise from a square wave generator with the SNR under program control.

   b. Field noise digitally added to speech.

      (1) Speech plus noise from office or helicopter environment with SNR under program control.

c. Field noise acoustically added to speech

(1) Live recordings made with portable cassette recorder.

(2) NSA recordings.

2. Utility programs have been developed for measuring signal energy and controlling signal-to-noise ratio, SNR.

E. Future plans

1. To receive and digitize any additional audio tapes provided by NRL needed to evaluate noise suppression algorithms.

SECTION III

Characterization of the Performance of Current LPC Speech Analysis Methods Applied to Noisy Speech

Introduction

This section considers two alternative but complementary methods for characterizing the effect of noise on the LPC analysis of speech. First, time and frequency relationships between the speech, noise and noisy speech are computed and compared. Second, LPC parameter variations due to additive noise are investigated.

Time-Frequency Relations

A. Objectives

1. Investigate modifications to the short time speech spectrum and corresponding all-pole spectrum caused by additive noise.
2. Determine to what extent the speech and noise are correlated during vocoder analysis time segments.
3. Examine the temporal modifications to the synthetic speech resulting from the analysis of noisy speech.

B. Approach

1. The intelligibility and quality of vocoder speech depends directly on how closly the all-pole spectrum

matches the noise free spectrum. To determine how noise modifies or distorts the spectral bit from the noise free case, a standard LPC analysis was applied to both speech and speech plus wide band Gaussian noise at specified signal-to-noise ratios.

2. Classical methods for suppressing noise using linear filtering usually make the assumption that the desired signal is uncorrelated with additive broadband Gaussian noise. This simplifies the analysis, since now crosscorrelations between speech and noise are set to zero. To determine whether this assumption remains valid over the short analysis periods encountered in speech processing the short time crosscorrelations and autocorrelations between speech and broadband Gaussian noise are computed and compared. If the crosscorrelations are not negligible then they must be accounted for when analyzing noisy speech.

3. Basic to any investigation in speech analysis is ability to interactively display and listen to the synthetic speech derived from the analysis. Therefore, synthetic speech was generated using clean and noisy speech at specified signal-to-noise levels. Critical headphone listening tests were then made to judge subjective changes in quality and intelligibility due to the addition of noise.

C. Tasks

1. Development of a General Purpose Waveform Display Program

a. Program Name: DSPLAY

b. Program Author: Dennis Pulsipher

c. Motivation: In order to evaluate the performance of LPC analysis, it is essential that time waveforms and their spectra be available for interactive display and playback. A flexible, interactive graphics and audio playback program was developed to accomplish this task.

d. Program Features: The following is a copy of the options available for displaying data.

```
.RUN DSPLAY[21,21]

DISPLAY PROGRAM
EQUAL SIZE BUFFERS FOR TRACKS 1 AND 2
MAXIMUM LOG LENGTH: 8192
NOVEMBER 4, 1976  VER. 43

PLOT : >?
 TIME WAVEFORM
 MAGNITUDE OF FREQUENCY RESPONSE
 LOG-MAGNITUDE OF FREQUENCY RESPONSE SQUARED
 0 (PHASE OF FREQUENCY RESPONSE)
 NOTHING--SET STEREO TRACK
 EXPANDED VIEW
 OUTPUT TO DECTAPE OF TIME WAVEFORM
 BRAND NEW TIME WAVEFORM FROM DECTAPE
 CHANGED TIME WAVEFORM LENGTH
 UNFORMATTED TIME WAVEFORM FROM DISK
 SUCCEEDING SEGMENT FROM FILE
 FOLLOWING SEGMENT
 WINDOWED TIME WAVEFORM
 MODIFIED HAMMING WINDOWED RESPONSE
 TITLE GRAPH
 PLAY DISPLAY
 +
```

```
-
ADVANCE (OR BACKSPACE) RECORDS
DESCRIBE POSITION
# SUM OF SQUARES
STEREO SWITCH
SAMPLING FREQUENCY
B-WEIGHTING
MISCELLANY
GRAY & MARKEL DISTANCE MEASURES
SPECTRAL ESTIMATE (LPC)
MISCELLANY >>?
 LABELS
 NO LABELS
 DUAL GRID
 GRID
 LOG
 LINEAR
 MULTIPLE PLOT
 ONE PLOT
 IMPULSE/FREQUENCY
 TRACK1/TRACK2
 Y SCALE
 YOFF
 X SCALE
 XOFF
 VLOG
 VLIN
 ALL LABELS
 PRIMARY LABELS
 BIAS
 0 BIAS
 SET #
 CLEAR #
 PLOTS
 BLANKS (NO PLOTS)
 RETURN
 /
 RETURN
```

e. Display examples generated by this program are presented in the section on results.

2. Development of Spectral and Correlation Display Program Needed to Compare LPC Analysis

a. Program name: CMPARE

b. Program author: William Done

c. Program Features: The following is a

description of program CMPARE:

Calculation of linear prediction coefficients for speech in the presence of noise requires the development of software for simplified analysis of new noise cancellation procedures. The software should also provide linear prediction coefficients based on contaminated and uncontaminated speech as a standard of comparison for the algorithm being evaluated. A graphics system, based on a linear prediction vocoder, was developed to perform the following tasks:

(1) Calculate Mode 1 coefficients $a_1(i)$ from $s(k)$, the uncontaminated speech;

(2) Calculate Mode 2 coefficients $a_2(i)$ from $x(k) = s(k) + n(k)$, the contaminated speech;

(3) Calculate Mode 3 coefficients $a_3(i)$ from $\hat{S}(k)$ using the algorithm being tested.

Thus, the $a_1(i)$ represent LPC coefficients obtained from high quality speech, while the $a_2(i)$ are coefficients from noisy speech, and represent the quality possible in LPC if no noise removal is done.

The Mode 3 coefficients $a_3(i)$ are determined by the noise suppression algorithm being evaluated. This mode can be changed by changing one subroutine in the software. Associated with each Mode 3 system is a graphics routine

which allows important sequences of that algorithm to be displayed. The computer sense switches are used to select and load various arrays into the graphics software for plotting and determination of spectra. An example of a typical menu of arrays available for loading is listed below.

```
   MENU FOR LOADING

  4   PITCH PROFILE
  5   MODE 1 LOSS FUNCTION
  6   MODE 2 LOSS FUNCTION
  7   MODE 3 LOSS FUNCTION
  8   MODE 4 LOSS FUNCTION
  9   A1 PREDICTORS
 10   A2 PREDICTORS
 11   A3 PREDICTORS
 12   S(K)
 13   X(K)
 14   N(K)
 15   RSS(K)
 16   RXX(K)
 17   RNN(K)
 18   RXS(K)
 19   RSX(K)
 20   RNS(K)
 21   RXN(K)
 23   RNX(K)
 24   DISTANCE MEASURE
 25   DUMMY
 26   RSSHAT(I):    UNCORR.
 27   RSSCOR(K):    CORR.
 28   ACOR(K):    CORR. PRED

# COMMANDS ARE
CLEAR ARRAY                  - Zero plotting buffers
UNWINDOWED DFT               - Compute DFT of a sequence
WINDOWED DFT                 - Window sequence, compute DFT
LOAD DATA ARRAYS             - Load plotting buffers
DISPLAY                      - Enter display routine
AUTO- & CROSS-CORRELATIONS   - Compute those for s(k), x(k), n(k)
FLAG FOR HALT                - Set flag to stop at a specific frame
TRANSFORM TYPE               - Transforms to be magnitude or log magnitude
MENU FOR LOADING             - List menu above
ALL-POLE CROSS SPECTRA       - Compute all-pole spectra
QUIT PROGRAM                 - Exit graphics routine
```

Operations available in the graphics routine are also listed above with an explanation of their function. Below are listed the commands for the display programs.

```
 DISPLAY

NPLOTS =    3
>?
 GRID TYPE
 COMPLEXITY OF GRID
 SINGLE PLOT
 MULTIPLE PLOTS
 DISPLAY SIZE
 INTENSITY
 XMIN
 ABCISSA VALUES
 Y VALUES
 LIMITS ON Y
 PLOT
 HORIZONTAL LABEL
 VERTICAL LABEL
2 PLOTS
RETURN


2 PLOTS >>?
 Y VALUES:
 ABSCISSA VALUES:
 XMIN:
 DISPLAY SIZE:
 COMPLEXITY:
 LIMITS ON Y:
 PLOT:
 MULTIPLE PLOTS:
 SINGLE PLOTS:
 LABEL
 SWAP
 RETURN
 CLEAR
```

The system as described above is versatile in allowing the researcher to bring a new algorithm into operation quickly, with the facility of being able to generate spectra of processed sequences.

3. Comparison of the LPC Spectral Analysis of Clean and Noisy Speech

   a. Spectral comparisons: Using the utility programs described above, wide band Gaussian noise was scaled and added to a clean speech file resulting in an average signal-to-noise ratio of 0dB. Using this data DFT spectra and all-pole spectra of the speech, noise and noisy speech were computed and made available for display. Examples of these spectral comparisons are given in the next section.

4. Synthesis Speech Comparisons: Using the 0dB data base, LPC synthesis speech was generated, displayed and recorded. To eliminate differences in quality or intelligibility due to pitch and voicing differences between clean and noisy speech, the excitation parameters were computed using the clean speech file. This required modifying the LPC vocoder program to now accept two data files simultaneously: clean speech and noise. Pitch and voicing decisions were made from the clean text, LPC parameters and gain were computed from the sum of speech and noise. Examples of the synthesis differences are given in the next section.

5. Crosscorrelation Comparisons: Using the 0dB data base, the autocorrelations and crosscorrelations

required for LPC analysis of noisy speech were computed and made available for display. Of specific interest in this investigation was the determination of how correlated the speech and noise waveforms are within the given short time analysis frame used by the vocoder analyzer. The correlations considered are:

Model $x(m) = S(m) + n(m)$

Clean speech Correlations: $$R_{ss}(k) = \sum_{m=0}^{N-1} S(m)S(m+k)$$

Noisy speech Correlations: $$R_{xx}(k) = \sum_{m=0}^{N-1} x(m)x(m+k)$$

Speech-Noise Crosscorrelations: $$R_{sn}(k) = \sum_{m=0}^{N-1} S(m)n(m+k)$$

Assuming the additive noise model, the clean speech autocorrelations needed to solve for predictor coefficients are given by

$$R_{ss}(m) = R_{xx}(m) - R_{xn}(m) - R_{nx}(m) + R_{nn}(m)$$

or in terms of the crosscorrelations between the speech s(m) and noise n(m):

$$R_{ss}(m) = R_{xx}(m) - R_{sn}(m) - R_{ns}(m) - R_{nn}(m)$$

Classical linear filtering analysis methods such as Wiener filtering and Kalman filtering assume that signal and noise are uncorrelated and therefore, that $R_{sn}(m)$ and $R_{ns}(m)$ will average to zero. Although this may be true for long time averages, the nonstationarity of the speech prohibits averaging longer than 20 or 30 milliseconds. For this short time interval the cross-terms must be examined and included in the analysis if their values are substantial. The next section gives representative examples of the cross-terms.

D. Results

1. Introduction: This section presents a number of representative examples of time and spectral relations between speech, noise and their sum. Based upon a preliminary analysis of the office and helicopter environments, the primary noise component present is broadband white Gaussian noise. Therefore, it was decided to use this type of signal contamination to determine how LPC analysis degrades. In the following examples digitized Gaussian noise was scaled and added to clean speech to give an average signal-to-noise ratio of 0dB.

2. Spectral Distortion due to Additive Noise

   a. Major differences between the all-pole spectra of clean and noisy speech include:

(1) Loss of low energy formant information.

(2) Shifted formant frequencies

(3) Widened formant bandwidths

(4) Overall decrease of spectral dynamic range

b. The following figures demonstrate these effects clearly.

(1) Figure III.1 presents a dual plot of clean speech s(k) and the noisy speech x(k).

(2) Figure III.2 presents a dual plot of (in the upper trace) the corresponding spectrum and all-pole spectrum of s(k), and (in the lower trace) the spectrum and all-pole spectrum of x(k).

(3) Figure III.3 presents multiple plots in the top trace of the spectrum of the noisy speech x(k) its all-pole LPC spectrum, XHAT(k) and the all-pole LPC spectrum of the clean speech, SHAT(k). In the bottom trace is the spectrum of the noise N(k) which was added to s(k).

c. Comments: Figure III.3 clearly demonstrates how the all-pole spectrum of x(k) differs from that of s(k). Since the first formant has maximum energy its approximation is not noticably modified. However, the second formant of XHAT(k) is both shifted and its

bandwidth approximately tripled with respect to the second formant of SHAT(k). The third formant of XHAT corresponds to a high energy peak of N(k) occurring at about 2200 HZ, rather than the actual peak at 1750 HZ. Finally, the spectral dynamic range has decreased from about 55dB to about 20dB.

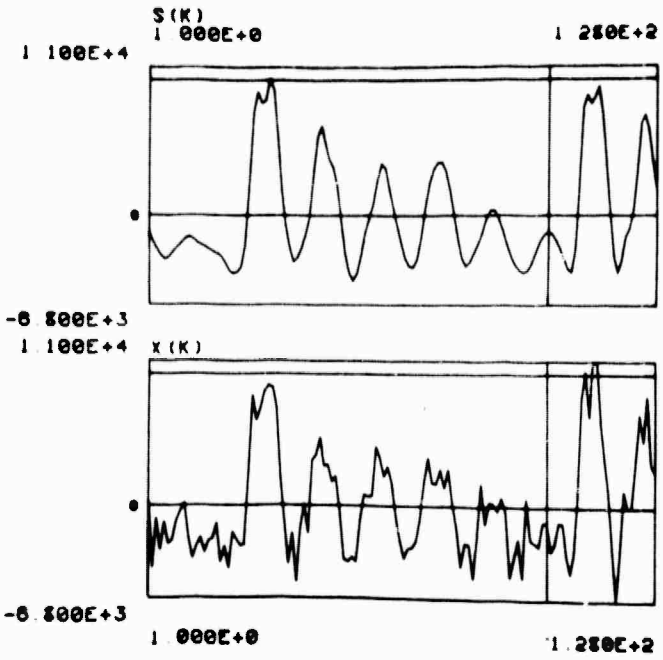


Figure III.1
Speech, S(k) and Speech plus Noise, X(k)

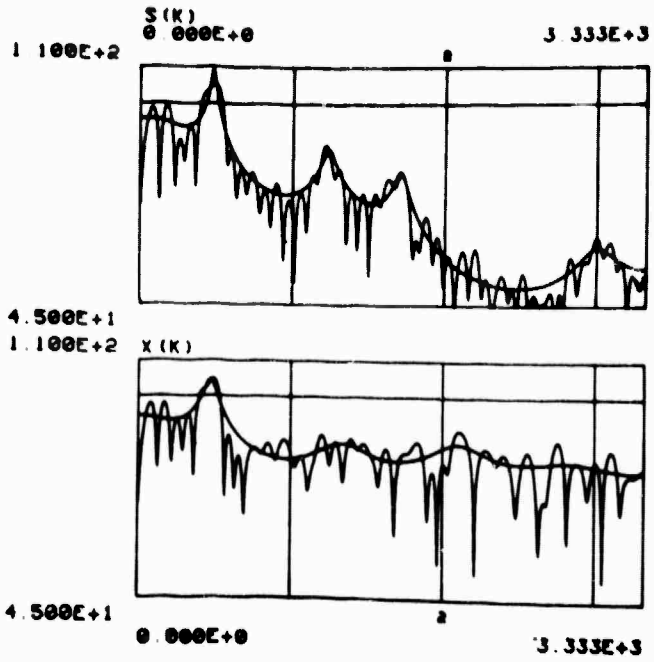


Figure III.2
Spectra and All-pole Spectra for the Speech, S(k)
Speech Plus Noise, X(k)

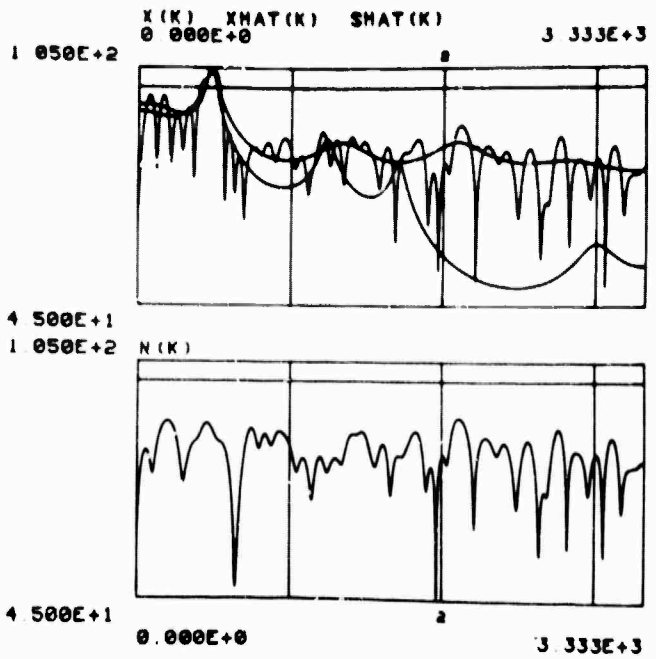


Figure III.3

Spectra of Noisy Speech X(k), All-pole Estimate XHAT(k), All-pole Estimate of Clean Speech SHAT(k), and Additive Noise, N(k).

3. Modifications to the synthetic speech waveform

a. The corresponding temporal distortions resulting from LPC spectral analysis of noisy speech are:

(1) Absence of "ringing" due to loss of low energy formants.

(2) Increase in "buzzy" quality due to spectral flattening.

(3) Increase in pitch and voicing errors.

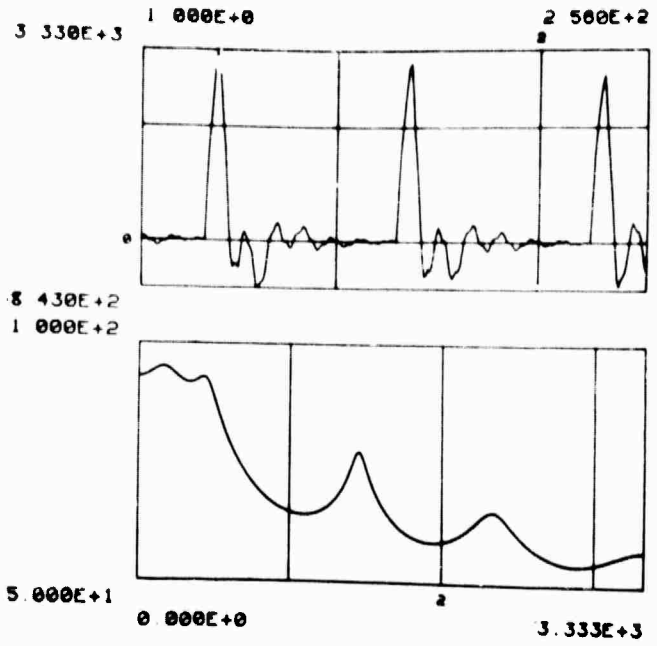
b. The following figures demonstrate these effects:

(1) Figure III.4 (a) presents the synthetic waveform and its all-pole spectrum using clean speech s(k).

(2) Figure III.4 (b) presents the synthetic waveform and its all-pole spectrum using noisy speech (0dB SNR).

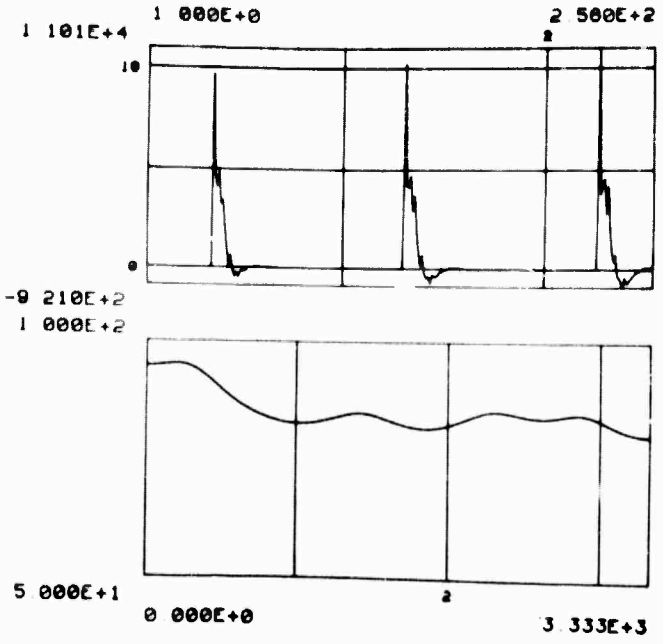
(3) Note: these examples are taken from another time window of the same speaker.

c. Comments: The severly overdamped character of the noisy speech synthesis clearly demonstrates why it will sound more buzzy and be less intelligible than the noise-free synthesis.



(a)

Clear Speech Synthesis and Spectrum



(b)

Noisy Speech Synthesis and Spectrum (0dB)

Figure III.4

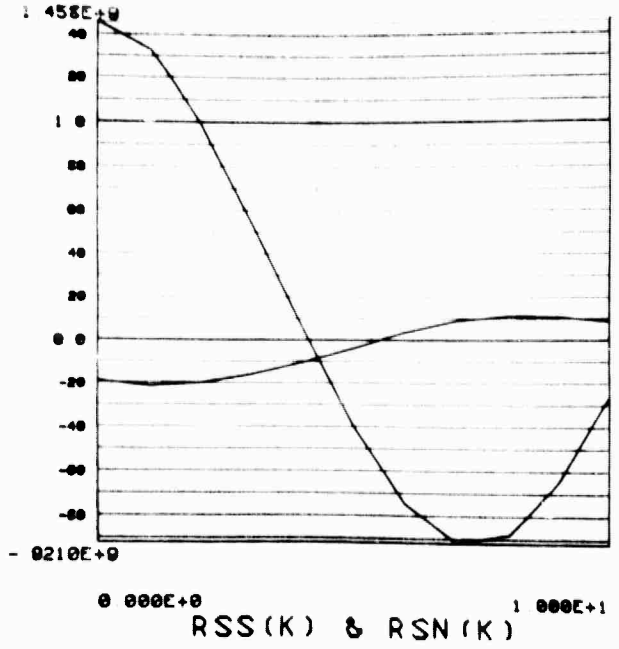


Figure III.5

Autocorrelation of Clean Speech RSS(k) and Crosscorrelation between Speech and Noise RSN(k)

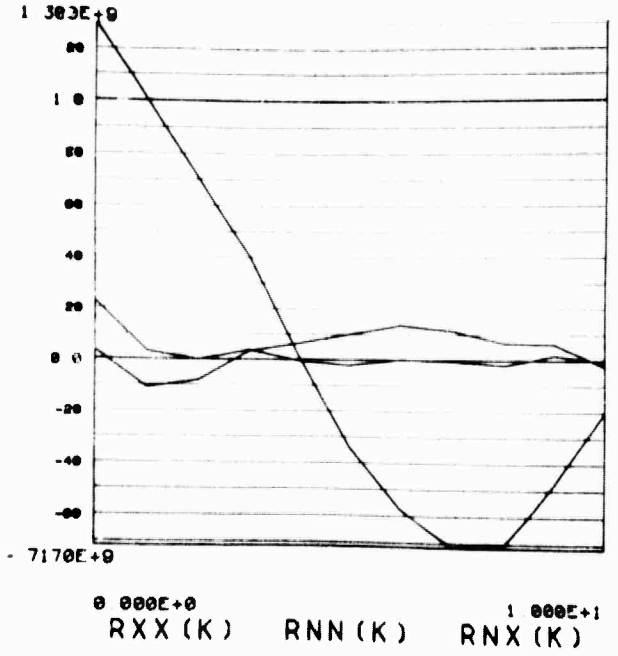


Figure III.6

Autocorrelation of Noisy Speech RXX(k) Noise RNN(k) and Crosscorrelation RNS(k)

4. Short time correlations between speech, noise and noisy speech.

   a. Using data corresponding to an analysis windowlength of 19.2 ms, the autocorrelations and crosscorrelations between speech, noise and noisy speech were computed and displayed.

   b. Figure III.5 presents a dual plot of a representative example of the short time autocorrelation of the clean speech $R_{ss}(k)$ and the short time crosscorrelation of the clean speech and noise $R_{ns}(k)$. (Data used is shown in Figure III.1)

      (1) Where

$$R_{ss}(k) = \sum_{m=1}^{128} S(m)S(m+k) \qquad k = 0, 1, \ldots, 10$$

$$R_{sn}(k) = \sum_{m=1}^{128} S(m)n(m+k) \qquad k = 0, 1, \ldots, 10$$

   c. Figure III.6 presents a triple plot for the same data of the autocorrelation of the noisy speech RXX(k), the autocorrelation of the noise RNN(k) and the crosscorrelation between the noisy speech and the noise RNX(k).

   d. Comments: From these figures it is clearly evident that for the short averaging intervals imposed on by the brief stationarity of the

speech signal, that the cross-terms are not given a chance to average to zero, and thus they cannot be ignored.

LPC Parameter Variations Due To Additive Noise

A. Objectives

1. Compare LPC analysis parameter variations versus Signal Contamination

B. Approach

1. Using the laboratory noise data base described in Section II noisy speech files with specified average signal-to-noise ratios were created.
2. Using these calibrated noisy speech file time histories of the LPC analysis parameters were generated and saved.
3. A general purpose parameter comparison program was written to examine, display, and summarize parameter variations versus signal-to-noise levels.
4. Initially the parameters were computed for the non-distorted speech for each analysis frame and stored as a reference parameter file. The parameter computation was then repeated on the degraded speech. The resulting parameter files were then compared on a frame by frame basis.
5. As noise suppression algorithms are developed, their ability to improve vocoder performance will be

empirically measured by comparing the analysis parameters from the noise cancelled process with those generated from clean speech.

C. Tasks

1. Development of a LPC analysis parameter generator program.

   a. Program name: ANALYS.SAV

   b. Program author: R. Frost

   c. Program Description: ANALYS.SAV is an adaptation of S. Boll's vocoder program. It writes out on disk the energy, error energy, pitch, voicing decision, LPC predictor coefficients, reflection coefficients, and speech autocorrelations at each analysis frame. The number of poles is variable, and is indicated by the variable k in the first data instruction.

2. Development of a parameter comparison program.

   a. Program name: NUCMPR.SAV

   b. Program author: R. Frost

   c. Program Description: NUCMPR.SAV (FOR NEW COMPARE) is the comparison program. It reads two parameter files created by ANALYS.SAV, which maybe up to 20000 points long. For a 12 pole LPC analysis, this corresponds to about 64000 speech data points. Five basic options

are available: (1) listing of the parameters on a frame by frame basis for the two files, or (2) viewing an overall comparison of the pitch characteristics of the two files, including plots of the pitch histories, the number of cross pitch errors (>10ms), voiced to unvoiced errors, unvoiced to voiced errors, and the mean and standard deviation of the fine pitch errors, (3) plots of the energy in each speech file, (4) plots of the error energy in each file, and (5) plots of both the maximum and minimum distances between the files, as described by Gray and Markel, "Distance Measure for Speech Processing", IEEE-ASSP, Vol 24, No. 5, pp. 380-391. Their approach is to define a metric based on the rms log spectral distance. This distance can then be computationally estimated in an efficient manner by computation of an upper and lower bound, the "cosh" and "cepstral" approximations, respectively.

These separate functions are obtained by typing the appropriate command after the command generator herald, which is && .

3. Results using program NUCMPR:

a. Our experience with this approach is that

these measures are consistent, and are evaluated reasonably quickly. As an example, an utterance was corrupted by adding various amounts of noise. The distances between the various versions were computed, and are tabulated below. In each case the distance is measured from the utterance having a SNR of 40dB.

SNR	Cosh Distance Measure	Cepstral Distance Measure
40dB	mean = 0dB σ = 0dB	mean = 0dB σ = 0dB
30dB	mean = 2.50dB σ = 1.14dB	mean = 2.37dB σ = 1.07dB
20dB	mean = 5.35dB σ = 2.04dB	mean = 4.50dB σ = 1.69dB
10dB	mean = 7.74dB σ = 2.61dB	mean = 6.15dB σ = 2.19dB
0dB	mean = 9.42dB σ = 3.33dB	mean = 7.33dB σ = 2.88dB
-10dB	mean =10.70dB σ = 4.39dB	mean = 8.14dB σ = 3.64dB

In general, our experience is consistent with the conjecture of Gray and Markel that distances of less than about 2dB are difficult to perceive, while greater distances are quite noticeable, and become increasingly offensive.

b. As noted above, the time histories for various vocoder parameters can be saved and plotted for comparison. As an example of the outputs

available, time histories of the pitch period, energy, error energy and Gray and Markel spectral distances were computed for the clean, noisy and processed speech used in the dual input adaptive noise cancelling experiment described in Section V . The data presented represents analysis parameters from 248 analysis frames.

For each figure, part (a) compares the parameter histories of clean speech versus processed speech and part (b) compares the parameter histories of the noisy speech versus the processed speech. The time histories include:

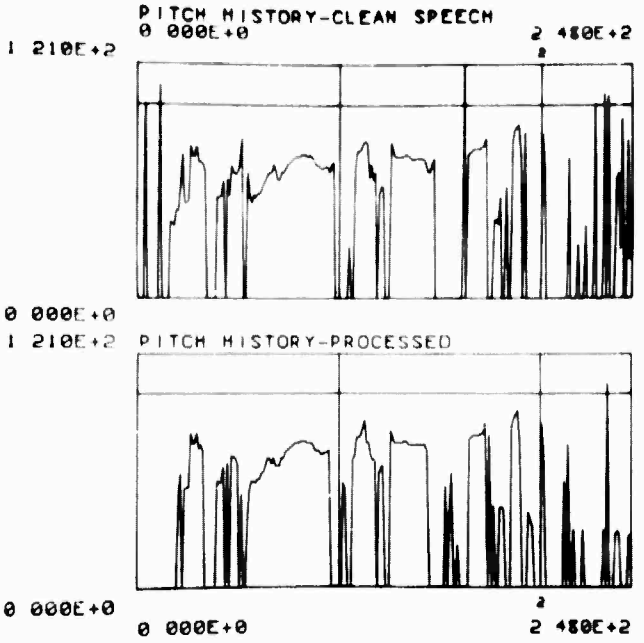
(1) Figure III.7 Pitch and Voicing (unvoiced equals zero)

(2) Figure III.8 Signal energy

(3) Figure III.9 LPC Prediction Error Energy

(4) Figure III.10 Maximum and Minimum Gray and Markel Spectral Distances

(a)



(b)

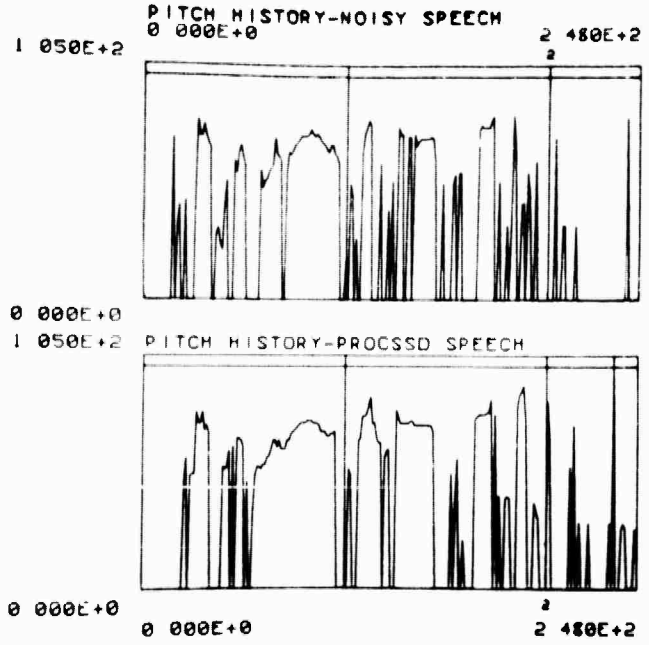
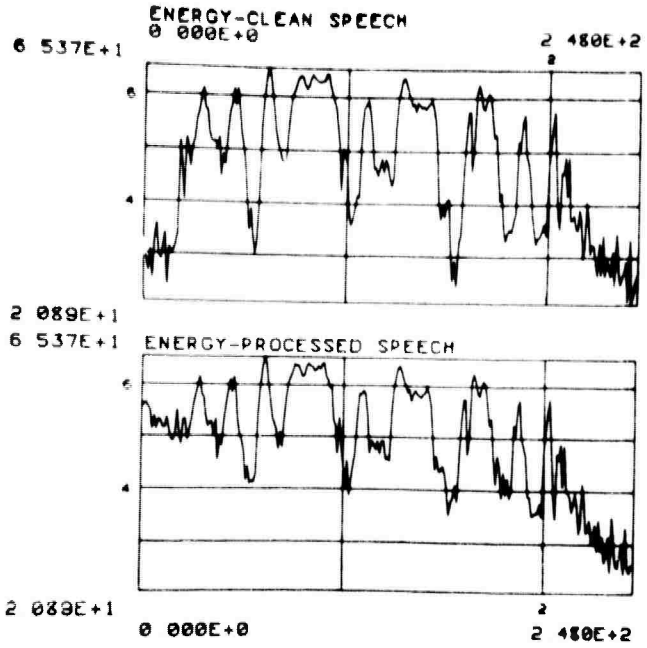


Figure III.7
Time Histories of Pitch and Voicing

(a)



(b)

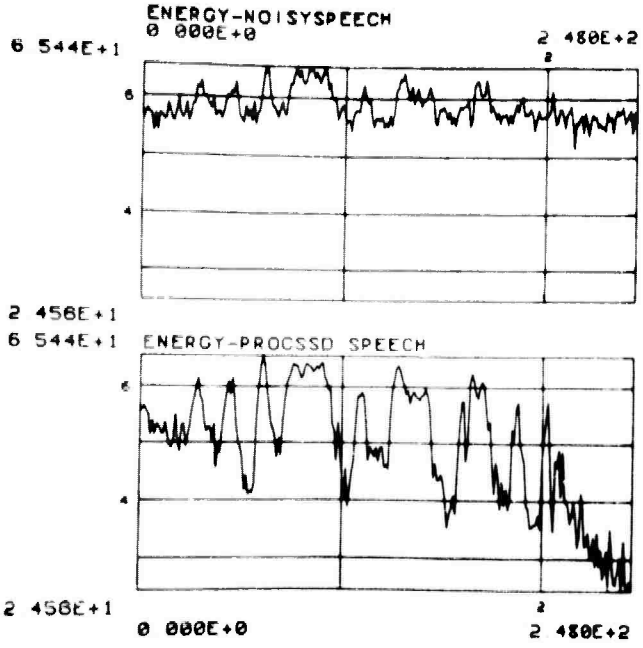
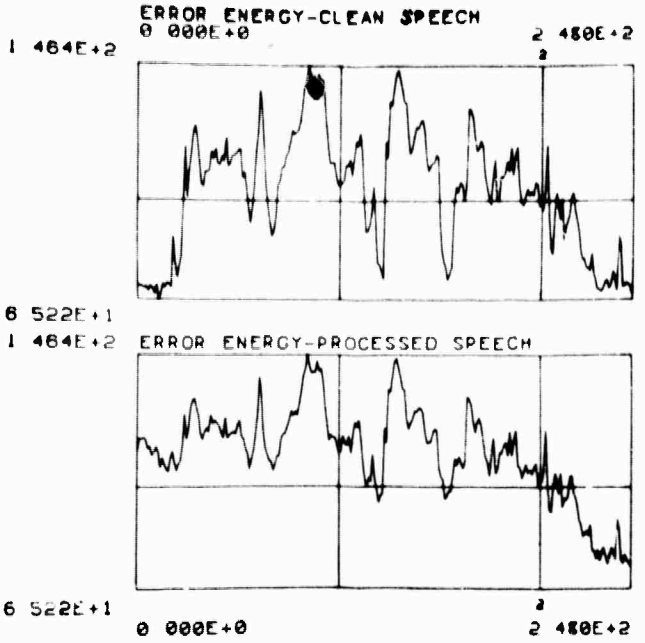


Figure III.8
Time Histories of Signal Energy

(a)



(b)

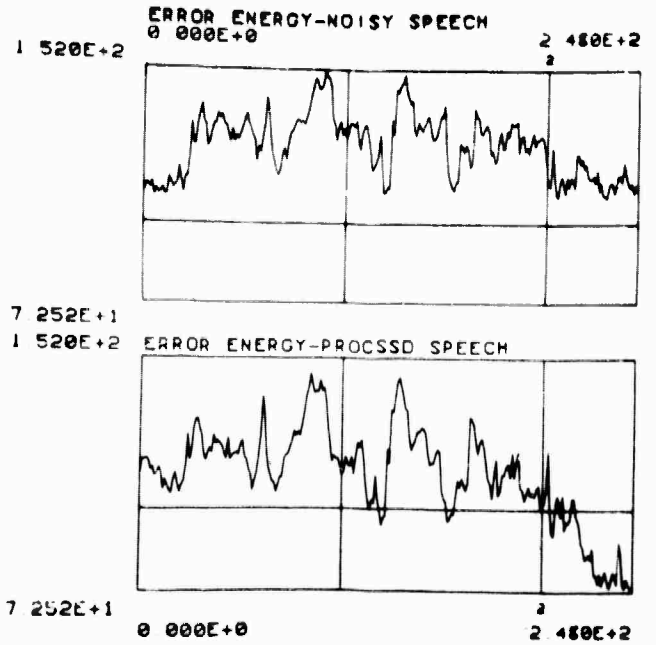
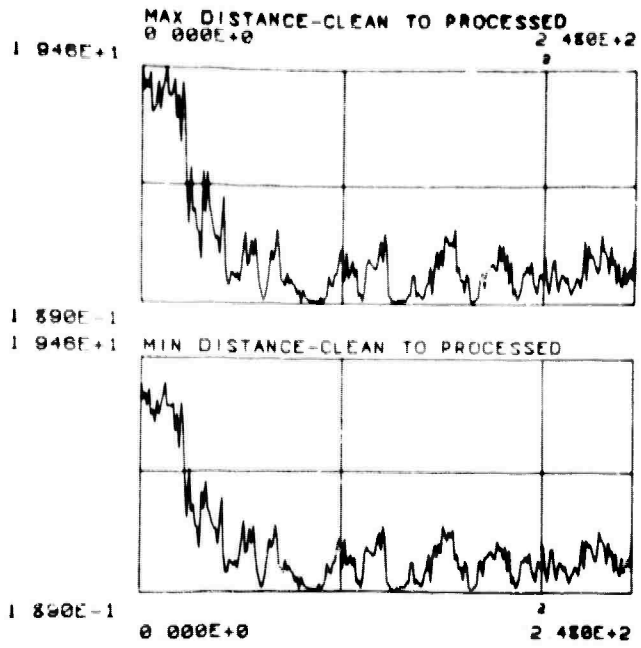


Figure III 9

Time Histories of LPC Prediction Error Energy

(a)



(b)

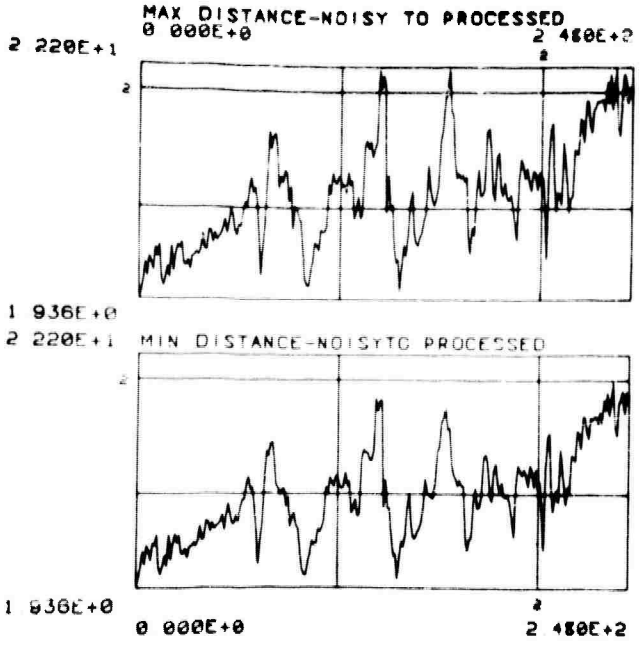


Figure III.10
Time Histories of Gray and Markel Spectral Distances

SECTION IV

An Integrated Noise Suppression-Speech Analysis Algorithm

Predictive Noise Cancellation

A. Preface

This section describes a preliminary investigation of a method for noise suppression where the analysis autocorrelations are modified to explicitly account for additive noise present on the speech waveform. The method, Predictive Noise Cancellation, gets its name from the fact that an estimate of the current noise is adaptively predicted from long term noise statistics. A description of the method is provided in the accompaning paper, "Improving Linear Prediction Analysis of Noisy Speech by Predictive Noise Cancellation", presented at the 1977 International Conference on Acoustics, Speech and Signal Processing, Hartford, Connecticut. The objectives, approach, tasks, and accompanying theory are given in the paper. Results and future research efforts implied from this study are listed below.

B. Results

1. Advantages of Predictive Noise Cancellation.

   a. The method uses procedures which are currently

available in real time LPC Vocoders.

(1) Autocorrelations and convolutions

(2) Levinson recursions

(3) All-pole synthesis

b. The method results in a stable synthesis.

(1) The autocorrelations can be guaranteed to be positive definite.

c. Background noise energy is reduced from 10dB to 20dB depending upon the noise environment.

2. Disadvantages of Predictive Noise Cancellation.

a. The method is dependent upon the phase of the signals processed since crosscorrelations are used.

(1) This required converting all signals to minimum phase realizations.

b. The noise-signal correlation filter estimate (Wiener Filter) was corrupted when speech was present.

(1) The estimator requires the crosscorrelation $R_{\bar{n}n}(k)$ between the current noise n(k) and the average noise $\bar{n}(k)$. This correlation term is given by:

$$R_{\bar{n}n}(k) = R_{\bar{n}x}(k) - R_{\bar{n}s}(k)$$

(2) When $R_{\bar{n}s}(k)$ is non-zero (speech present),

$\bar{R}_{nx}(k)$ becomes a poor estimate of $\bar{R}_{nn}(k)$.

C. Future Research

Based upon the inadequacies of the Predictive Noise Cancellation model, a frequency domain spectral averaging technique is currently being eveloped. This method retains the advantages of PNC (ie LPC algorithms, stable synthesis, and about 15dB noise suppression) but avoids the disadvantages. Results will be presented in the next Semi-annual Technical Report.

IMPROVING LINEAR PREDICTION ANALYSIS OF NOISY SPEECH BY PREDICTIVE NOISE CANCELLATION

Steven F. Boll

Computer Science Department
University of Utah
Salt Lake City, Utah 84112

Abstract

The analysis of speech using Linear Prediction is reformulated to account for the presence of acoustically added noise and a technique is presented for reducing its effect on parameter estimation. The method, called Predictive Noise Cancellation (PNC), modifies the noisy speech autocorrelations using an estimate of present background noise which is adaptively updated from an average all-pole noise spectrum. The all-pole noise spectrum is calculated by averaging autocorrelations during non-speech activity. The method uses procedures which are already available to the LPC analyzer, and thus is well suited for real time analysis of noisy speech. Preliminary results show signal to noise improvements on the order of 10 to 20 db.

Introduction

As noise is acoustically added to speech, the resulting intelligibility and quality of the LPC synthesis degrades [1], [1]. This paper presents a technique which accounts for the noise present and modifies the noisy speech autocorrelations in order to suppress it. The method is based upon the simple observation that if $x(k)=s(k)+n(k)$, where $s(k)$ is clean speech, $n(k)$ is the added noise, and $x(k)$ their sum, and if the noise signal $n(k)$ were known exactly, then the desired speech autocorrelations, $R_{ss}(m)$ can be recovered from the noisy speech, $x(k)$ by computing:

$$R_{ss}(m) = R_{xx}(m) - R_{xn}(m) - R_{nx}(m) + R_{nn}(m) \qquad (1)$$

where

$$R_{xn}(m) = \sum_k x(k)n(k+m) = R_{nx}(-m)$$

$$R_{xx}(m) = \sum_k x(k)x(k+m)$$

$$R_{ss}(m) = \sum_k s(k)s(k+m)$$

Of course the noise is not known within any given analysis frame and must be approximated. A method for estimating it is the subject of this paper. Once an estimate for the local noise component is determined, Equation (1) can be used to calculate the autocorrelations of the estimated speech spectrum from which the LPC parameters can be obtained.

Constraints

Since the noise cancellation is to be integrated into the LPC analysis, it was decided that the estimation of the present noise component be done using algorithms already available to the LPC analyzer. In addition, noise characterization and estimation should depend only upon the actual background environment as recorded by the microphone.

Plan

To satisfy these constraints the noise environment is modeled by an all-pole spectrum. It is estimated by averaging autocorrelations during an initial period of non-speech activity. These averaged noise autocorrelations are then used to estimate the present frame noise component. The local noise component is estimated by convolving the average noise autocorrelations with a correlation filter whose impulse response is estimated for each frame to minimize the mean square error between the average noise and the local signal. Thus the method can be described as that of adaptively filtering past noise to approximate present noise.

Method

There are four phases to the process of Predictive Noise Cancellation. They are: (1) estimation of average background noise using LPC; (2) estimation of noise-signal correlation filter; (3) modification of noisy speech autocorrelations; and (4) calculation of final LPC parameters.

Background Noise Estimation

During the startup or a calibration period when just background noise is recorded by the microphone, the first M+1 autocorrelations representing just noise are computed and averaged together. Define:

$$R_{\hat{n}\hat{n}}(m) = \frac{1}{N_c}\sum_{l=1}^{N_c} R_{xx}^{(l)}(m) \quad m=0,1,\ldots,M \tag{2}$$

where

$$R_{xx}^{(l)}(m) = \sum_{k=0}^{N-1} x(k)x(k+m),$$

is the mth autocorrelation during the lth frame to be averaged.

- $x(k)$ = noise signal ($s(k)=0$)
- N_c = number of frames to be averaged (normally 1/2 sec)
- M = order of all-pole noise spectrum (set to 10)

At the completion of the calibration period, predictor coefficients representing the noise $a_n(k)$ are computed using Levinson's recursion.

Finally since it will be necessary to compute crosscorrelation between the average noise $\hat{n}(k)$ and the noisy signal, $x(k)$, the first N values of the minimum phase impulse response $\hat{n}(k)$ defined from $a_n(k)$ are computed as:

$$\hat{n}(k) = -\sum_{i=1}^{M} a_n(i)\hat{n}(k-i) + G_{\hat{n}}\delta_{k,0} \tag{3}$$

$$k = 0,1,\ldots,N-1.$$

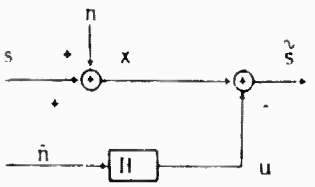
where

$$G_{\hat{n}}^2 = R_{\hat{n}\hat{n}}(0) + \sum_{i=1}^{M} a_M(i)R_{\hat{n}\hat{n}}(i)$$

N = analysis window length (Nominally 20 ms)

Noise-Signal Correlation Filter

A block diagram indicating the noise cancellation procedure is shown in Figure (1).



- s: speech
- n: noise
- $\hat{n}$: averaged noise
- $H(z) = \sum_{i=0}^{L} h(i)z^{-i}$: correlation filter
- x: noisy speech
- u: filtered average noise
- $\tilde{s}$: noise cancelled speech

Predictive Noise Cancellation Block Diagram
Figure 1

The purpose of $H(z)$ is to modify $\hat{n}(k)$ to approximate the noise $n(k)$ within the current analysis frame. The filter is estimated using a least square criterion. The tap parameters of $H(z)$ are estimated in order to minimize

$$\sum_{k} [\tilde{s}(k)]^2 = \sum_{k} [\hat{x}(k) - \sum_{i=0}^{L} h(i)\hat{n}(k-i)]^2 \tag{4}$$

Minimizing Equation (4) with respect to $h(i)$ results in a toeplitz system of linear equations:

$$\sum_{i=0}^{L} h(i)R_{\hat{n}\hat{n}}(i-j) = R_{\hat{n}\hat{x}}(j) \quad j=0,1,\ldots,L \tag{5}$$

where

$$R_{\hat{n}\hat{x}}(j) = \sum_{k=0}^{n-1} \hat{n}(k)\hat{x}(k+j)$$

$$\hat{x}(k) = -\sum_{i=1}^{M} a_x(i)\hat{x}(k-i) + G_x\delta_{k,0}$$

It was necessary to use the LPC minimum phase approximation $\hat{x}(k)$ to $x(k)$ since $\hat{n}(k)$ is an LPC minimum phase approximation. Note that $H(z)$ can be calculated using the two pass Levinson's recursion [2], [3].

After estimating $H(z)$ it is normalized to have a spectral average of unity by dividing each tap parameter $h(i)$ by $h(0)$. This normalization was included since the purpose of $H(z)$ is to shape the spectrum of $\hat{n}(k)$ but not to increase its total energy.

Autocorrelation Modification

Referring to Figure 2, the autocorrelation of the noise cancelled speech, $\tilde{s}(k)$ are given by

$$R_{\tilde{s}\tilde{s}}(m) = R_{\hat{x}\hat{x}}(m) - R_{\hat{x}u}(m) - R_{\hat{x}u}(-m) + R_{uu}(m) \tag{6}$$

$$m = 0,1,\ldots,M$$

It is not necessary to explicitly calculate $u(k)$ in order to obtain $R_{\hat{x}u}(m)$ and $R_{uu}(m)$. These correlation terms can be calculated from $R_{\hat{x}\hat{n}}(m)$ and $R_{\hat{n}\hat{n}}(m)$ as follows:

Since $u(k) = \sum_{i=0}^{L} h(i)u(k-i) \quad k=0,1,\ldots$ (7)

then

$$R_{\hat{x}u}(m) = \sum_{k=0}^{N-1} \hat{x}(k)u(k+m) = \sum_{k=0}^{N-1} \hat{x}(k) \sum_{i=0}^{L} h(i)\hat{n}(k+m-i) \tag{8}$$

or

$$R_{\hat{x}u}(m) = \sum_{i=0}^{L} h(i) \sum_{k=0}^{N-1} \hat{x}(k)\hat{n}(k+m-i) \tag{9}$$

In terms of $R_{\hat{x}\hat{n}}(m)$ we have

$$R_{\dot{x}u}(m) = \sum_{i=0}^{L} h(i) R_{\dot{x}\tilde{n}}(m-i) \qquad (10)$$

Likewise $R_{uu}(m)$ can be obtained from $R_{\tilde{n}\tilde{n}}(m)$ and h(i) as follows:

$$R_{uu}(m) = h(m) * h(-m) * R_{\tilde{n}\tilde{n}}(m) \qquad (11)$$

let

$$R_{hh}(m) = h(m) * h(-m) = \sum_{i=0}^{L} h(i)h(i+m) \qquad (12)$$

then

$$R_{uu}(m) = \sum_{i=-L}^{L} R_{hh}(i) R_{\tilde{n}\tilde{n}}(m-i) \quad m=0,1,\ldots,M \qquad (13)$$

LPC Parameter Calculation

Having calculated $R_{\dot{x}u}(m)$ and $R_{uu}(m)$, the autocorrelations $R_{\tilde{s}\tilde{s}}(m)$ of the noise cancelled speech can be computed using Equation (6). From these the LPC coefficients can be calculated using the Levinson's recursion. A stable filter will result since $R_{\tilde{s}\tilde{s}}(m)$ is positive definite.

Implementation and Results

The algorithm was inserted into an LPC vocoder simulation and tested on a data base consisting of three types of noisy speech. Type one was clean speech plus known amounts of guassian noise digitized from an analog noise generator. Type two was clean speech plus known amounts of noise recorded in a helicopter cockpit. Type three was speech recorded in a helicopter. Specifications for the vocoder simulation were as follows:

Sampling Frequency = 6.667 kHz
Analysis Window Length, N = 19.2 ms
Predictor Order, M = 10
Correlation Filter Order, L = 10
Initial Averaging Period, N_c = 0.5 sec.

Results

An audio tape demonstrating the results will be played. A coarse measure of signal to noise improvement can be calculated by comparing the energy before cancellation $R_{xx}(0)$ with the energy after cancellation $R_{\tilde{s}\tilde{s}}(0)$. An improvement on the order of 10 to 20 dh was observed for all types of noisy speech. Methods for measuring improvements in quality and intelligihility are currently being investigated.

Conclusion

An integrated system for noise cancellation coupled with LPC analysis has been presented. The method assumes that noise present during the current analysis frame can he estimated by filtering an all-pole average noise spectrum through an adaptively updated linear filter. The noisy speech autocorrelations are then modified to account for the noise estimate. The algorithm is currently being tested on a variety of noisy operating environments with preliminary results showing a signal to noise improvement of 10 to 20 db.

Acknowledgements

The author would like to thank William Done and Dennis Pulsipher for their assistance in this research.

This research is supported hy the Information Processing Techniques Branch of the Advanced Research Projects Agency.

References

1. B. Gold, "Robust Speech Processing", Technical Note 1976-6, Lincoln Laboratory, M.I.T., January 1976.

2. J. Markel and A. Gray, "On Autocorrelation Equations as Applied to Speech Analysis", IEEE Trans. on Audio and Elect., Vol. 21, No. 2, April 1973, p 69-79.

3. E. A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Holden Day, New York 1967, (Subroutine EUREKA).

4. M. Sambur and N. Jayant, "LPC Analysis/Synthesis from Speech Inputs Containing Quantizing Noise or Additive White Noise", IEEE Trans. on Acoust., Speech and Signal Processing, Vol. 24, No. 6, December 1976.

SECTION V

A PREPROCESSING NOISE CANCELLATION ALGORITHM: DUAL INPUT NOISE SUPPRESSION

DENNIS PULSIPHER

Introduction

The presence of noise in speech signals has long been annoying. Many techniques for reducing various types of noise have been described and implemented. These techniques have fallen mainly into two categories: direct linear filtering, and model fitting. Though many methods of deriving the filter to be used have been proposed, the filtering techniques have one thing in common; their attempt to improve the signal-to-noise ratio (SNR) is accomplished by attenuating those frequencies with poor SNR and giving emphasis to those with higher SNR, subject to certain other constraints.

Model fitting has been used to reduce noise by estimating a set of parameters which are then used to synthesize a signal estimate. Among the best examples of noise reduction by model fitting are vocoders, particularly the interactive homomorphic vocoder implemented by Neil Miller in 1973 [1].

New impetus has been given to research in the area of noise reduction by recent implementation of low-bandwidth digital transmission schemes, such as linear-predictive vocoders (LPC). In noisy environments such vocoders perform poorly.

Efforts to develop digital algorithms to minimize the noise-induced problems created by such environments as helicopters, airplanes, ships, and even offices have been intensified. Additional encouragement has been derived from the availability of digital processors capable of performing complex algorithms at real-time speeds.

It is in this setting we present a technique for using information obtained by making measurements of both a noisy signal and a signal containing only related noise to estimate a noise-free signal. We then present an algorithm for implementing the technique. A brief report on experiments performed to evaluate the technique, including data base generation and observations about the results then precedes the conclusion of the report. The observations and conclusions represent working ideas and should not be considered final at this time.

NOISE CANCELLATION

In December of 1975 Bernard Widrow proposed the use of Adaptive Noise Cancellation for the removal of noise from pilot-to-ground communication [2]. He also reported the results of an experiment using very stylized noise. Adaptive noise cancellation was not new, many applications had been found where it worked well. Among these were antenna side-lobe cancellation [3], data channel equalization [4], telephone channel echo cancellation [5] [6] [7], and noise reduction in electrocardiography [8]. Prior to this, however, no attempt that we are aware of was made to apply this technique to noise suppression in speech signals.

This noise cancellation technique differs significantly from classical techniques. Noise reduction is attempted by estimating the noise, then subtracting it from the noisy signal. Without using extreme care, this could result in an increase in noise power, so we should examine the mechanism by which a reduction is achieved.

If we are given the sum x of two mutually uncorrelated signals s, and N and a third signal V which is mutually uncorrelated with s, let us form a signal estimate s = x - u

where u is a linearly filtered version of V. (Figure V.1)

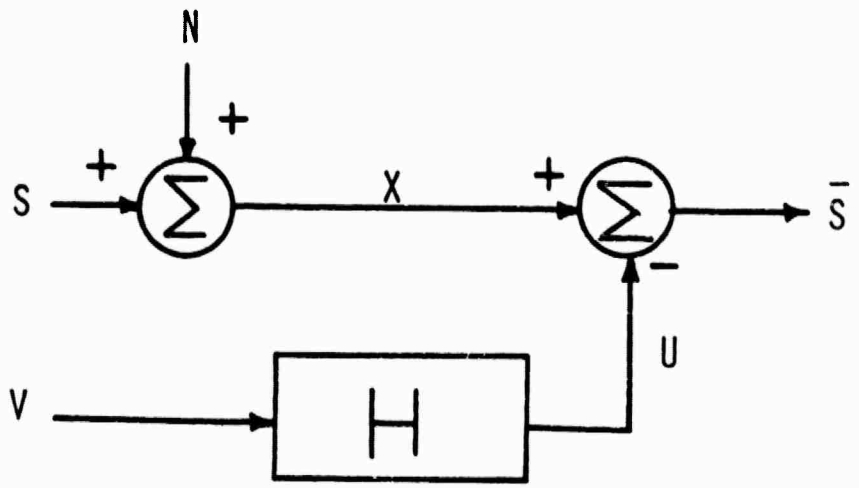


Figure V.1

Model for noise cancellation development

Then we have

1.) $\bar{s} = x - u = s + N - u$

so

$$\bar{s}^2 = s^2 + 2s(N-u) + (N-u)^2$$

and

$$E[\bar{s}^2] = E[s^2] + 2E[s(N-u)] + E[(N-u)^2]$$

Since s is uncorrelated with N and V

2.) $$E[\bar{s}^2] = E[s^2] + E[(N-u)^2]$$

This says that the average energy of $\bar{s}$ is the sum of the average energy in s and the average energy in (N-u). If we could minimize the energy in $\bar{s}$, (by adjustment of H), we would at the same time minimize the energy in (N-u), since

$E[s^2]$ is unaffected by changing H. That is, if we minimize the energy in $\bar{s}$, we have minimized the mean squared value of $(N - u)$, and consequently, the mean square value of $(\bar{s} - s)$, since by 1.) $\bar{s} - s = N - u$. Thus $\bar{s}$ is a mean squared estimate of s. The constraint, of course, is that our minimization of $E[\bar{s}^2]$ be accomplished with a u that is a linearly filtered version of V.

It remains for us to describe a model which could benefit from this analysis, and to present an algorithm for its implementation. The model we choose to assume for initial experimentation is one with a single signal source and a single noise source.

The noise V is recorded by a microphone placed so that the signal s does not reach it. The noise is also transmitted through a channel G and is recorded (at a second microphone) along with the signal s as the noisy signal x. (Figure V.2)

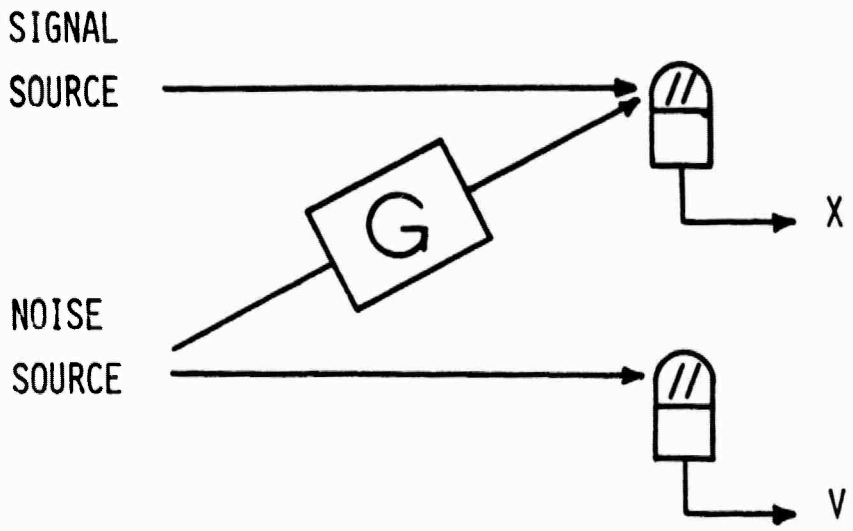


Figure V.2

Data generation model

If G can be approximated as a finite impulse response filter, clearly making H in Figure V.1 equal to G will result in a u equal to the noise in x. By subtracting u from x we are left with $\bar{s}$. In general, finding an optimal H is a problem equivalent to plant identification.

This analogy renews our hope that the technique may work well in extremely noisy environments, since H can be estimated best when x and V are closest to being linear filtered versions of each other. This happens when s is smallest relative to x. That is, we have the best chance of eliminating the noise in x when it is the noisiest - exactly when we need it the most for our application.

The model requires that we find an algorithm to best estimate H so that our mean squared estimate of s can be obtained.

THE ALGORITHM

The heart of the adaptive algorithm is the adaptive filter (channel) H_j. Through it pass the noise vectors V_j resulting in the noise estimate u_j. That is,

$$u_j = V_j^T H_j = H_j^T V_j$$

where V_j, H_j are L element vectors and j denotes the time at which they occur.

The estimate $\bar{s}_j$ is calculated by subtracting u_j from x_j.

$$\bar{s}_j = x_j - u_j = x_j - V_j^T H_j = x_j - H_j^T V_j$$

squaring yields

$$\bar{s}_j^2 = (x_j - u_j)^2 = x_j^2 - 2x_j V_j^T H_j + H_j^T V_j V_j^T H_j$$

and taking the expected value gives

$$E[\bar{s}_j^2] = E[x_j^2] - 2E[x_j V_j^T H_j] + E[H_j^T V_j V_j^T H_j].$$

Assuming a stationary filter H gives

$$E[\bar{s}_j^2] = E[x_j^2] - 2E[x_j V_j^T]H + H^T E[V_j V_j^T]H.$$

Defining

$$P \triangleq E[x_j V_j^T]$$

{ cross correlation between the noisy signal x_j (a scalar) and the noise reference V_j (a vector) }

and

$$R \triangleq E[V_j V_j^T]$$

[Reference input correlation matrix]

we have

$$E[\bar{s}_j^2] = E[x_j^2] - 2P^T H + H^T R H$$

which is a quadratic function of H, hence has a unique minimum H^*. By differentiating with respect to the elements of H we get

$$\nabla = -2P + 2RH.$$

Setting $\nabla = 0$ to find the minimum yields

$$H^* = R^{-1}P.$$

Since we are attempting to implement the process in real time we might use a steepest descent algorithm

$$H_{j+1} = H_j - \mu\nabla_j,$$

where the parameter μ controls convergence and stability. Unfortunately we do not have access to ∇_j, so we must be satisfied with a gradient estimate $\hat{\nabla}_j$. Widrow [2] has suggested the use of

$$\hat{\nabla}_j = -2\bar{s}_j V_j$$

which yields the algorithm:

$$H_{j+1} = H_j + 2\mu\bar{s}_j V_j$$

which is simple to implement.

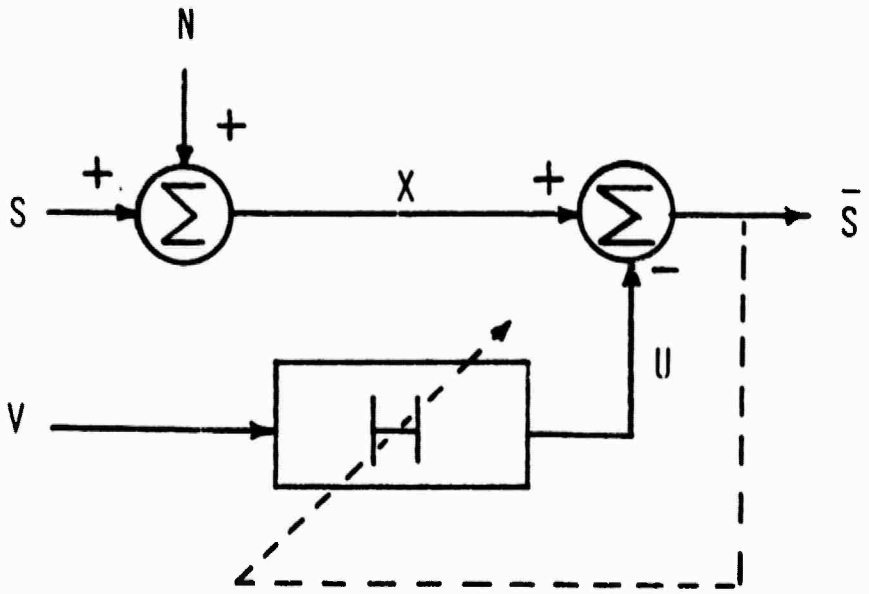


Figure V.3

Adaptive Model

Convergence of this estimate has been shown for V_j uncorrelated with V_k for k=j [9], provided μ is chosen small enough. Under the assumption of ergodicity, convergence has also been shown for special cases of correlated V_j [10], [11]. Researchers at Bell Laboratories have found the algorithm to be so robust that for echo suppression and channel equalization

$$\hat{\nabla}_j = -2\mathrm{SGN}[\bar{s}_j]V_j$$

where $\mathrm{SGN}[\cdot] \triangleq$ algebraic sign of $[\cdot]$

has been sufficiently accurate to produce convergence [5]. Others have proposed similar algorithms and proposed constant and time-varying μ's. [12] [13] [14] [15]. The algorithms asymptotic behavior, residual error, and nonstationary behavior in special cases have also been investigated elsewhere [2] [4] [13] [14] [15] [16].

THE EXPERIMENTS

In order to evaluate the potential of the adaptive noise cancellation algorithm as a technique for the suppression of noise in speech signals, several experiments were performed. Initially a data base of different types of noise was generated. Each of these noise sources was then passed through various known channels and the results used to augment the data base. This processed noise was then scaled and added to a speech signal. The resulting noisy signal and the original reference noise were then applied as inputs to the noise cancelling algorithm and the resulting filter estimates compared with the known channel responses. Similar experiments were performed on data collected in real situations.

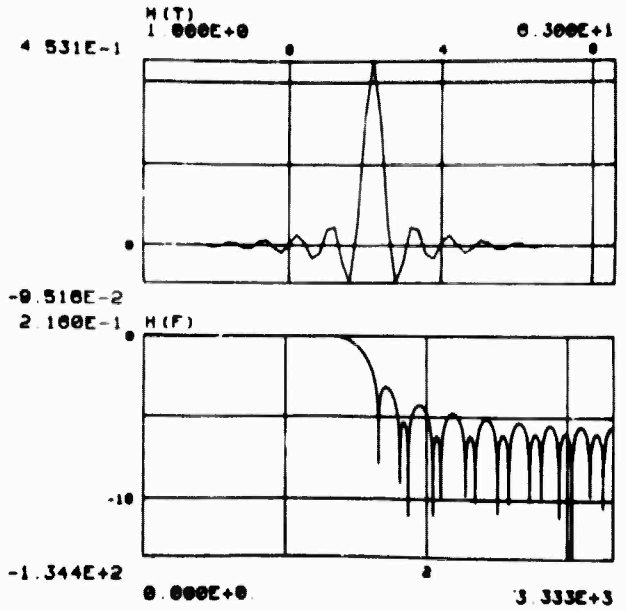
Providing a series of noise sources with varying characteristics, while keeping the base to a manageable size suitable for storage in a limited space was among the first

tasks to be undertaken. To be able to have unpredictable, yet stylized noise, we decided to digitize analog noise of three types. A Gaussian Noise Generator, a constant square wave, and a hand swept square wave were each low pass filtered to 3.2 KHZ and sampled at 6.67 KHZ. Approximately 12.3 sec. of each source was recorded.

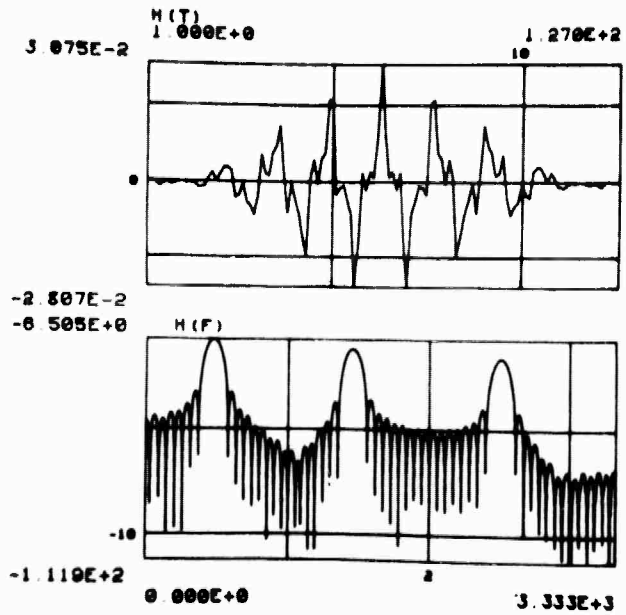
Three known channels were then used to process the previously digitized noise. A channel having a low pass cutoff at 1.5 KHZ, a channel with three narrow passbands at 500 HZ, 1500 HZ, and 2500 HZ, and a channel with a simple delay, were selected to represent a variety of possible channels. (Figure V.4).

Results of this processing were then measured by a spectral weighting algorithm which attempted to measure the energy in the signal, as perceived by a listener. Similar measurement was made of a speech signal, thus providing information making it possible to combine the two signals to yield a signal with a known SNR. The appropriate scaling was performed to provide signals with SNR of 0, 20 and 40 dB.

Data samples were also recorded in a quiet room, an office, and in a Bell Jet Ranger helicopter. Of course, in these cases the channel and SNR were not known, since two microphones were used to record the reference noise and the noisy signal simultaneously.



LOW PASS CHANNEL



THREE BAND PASS CHANNEL

Figure V.4

Known Channel Responses

RESULTS

The experiments performed yielded encouraging, but not decisive results. In the artificially contrived experiments it was found that each of the channels was estimated very closely for the cases where the reference noise was Gaussian, and had bands where the estimation was good and bands where it was bad for the other cases. This does not imply that significant noise reduction was not achieved, it simply implies that estimation of the channel was less accurate for these cases than for the Gaussian case. It is believed that this is caused because of the violation of two assumptions required for convergence to the optimal channel. The first of these is that successive noise vectors V_j are uncorrelated; for nearly periodic noise, this is clearly not true. The second is that an infinite sample size is available - this is also not true, though it appears to be the less important of the two problems.

Additional observations of interest include the fact that updating H_j on a point-by-point basis was much less irritating than updating less frequently. Since this practice forces successive V_j to be highly correlated, it may be interesting to note that correlation of successive V_j is not sufficient to prevent convergence of the channel estimate. The type of correlation is also very important.

It is also of significant interest to point out that a relatively fast adaptation rate seems to be preferred over a slow adaptation rate with a smaller residual error. That is, the open, echo-like quality induced by rapid adaptation is more pleasing than the loss of intelligibility in the first few words caused by slow adaptation. This is partly due to the slow disappearance of this echo with decreasing adaptation rate. Significantly, speech signals thus processed seem to work very well with LPC.

Other interesting results are that the signal estimate for the 0dB SNR signals are very similar to those for the 20dB SNR signals and the 40dB SNR signals. The change in the 40dB SNR signals should probably be classified as a slight degradation, while both other cases should be classified as major improvements.

Initial experiments on data recorded in actual settings have been less successful. While slight noise reductions have been observed, the failure of the channel estimate to converge to anything has spurred further experiments to determine the simple model's major deficiencies. Additional mutually uncorrelated noise at both inputs is an obvious source of error, but it is not known at this time how great the contribution of this source is. Other possible sources of error presently under investigation are multiple noise sources, non-linear channels, and the need for subsequent

processing of the signal estimate using conventic al techniques.

CONCLUSIONS

Many experiments have been performed and many are now being performed which suggest that the adaptive noise cancellation algorithm holds great promise for significant noise suppression in speech signals. While certain deficiencies have been observed for periodic noise, significant noise reduction has also been achieved. Though the signal estimates are sometimes excellent (see Figures V.5, V.6) a very slight degradation of the speech seems to be common. Perhaps most important is the realization that for real speech data the adaptive-noise cancellation algorithm has demonstrated a great potential for success but is not yet ready to claim a history of success.

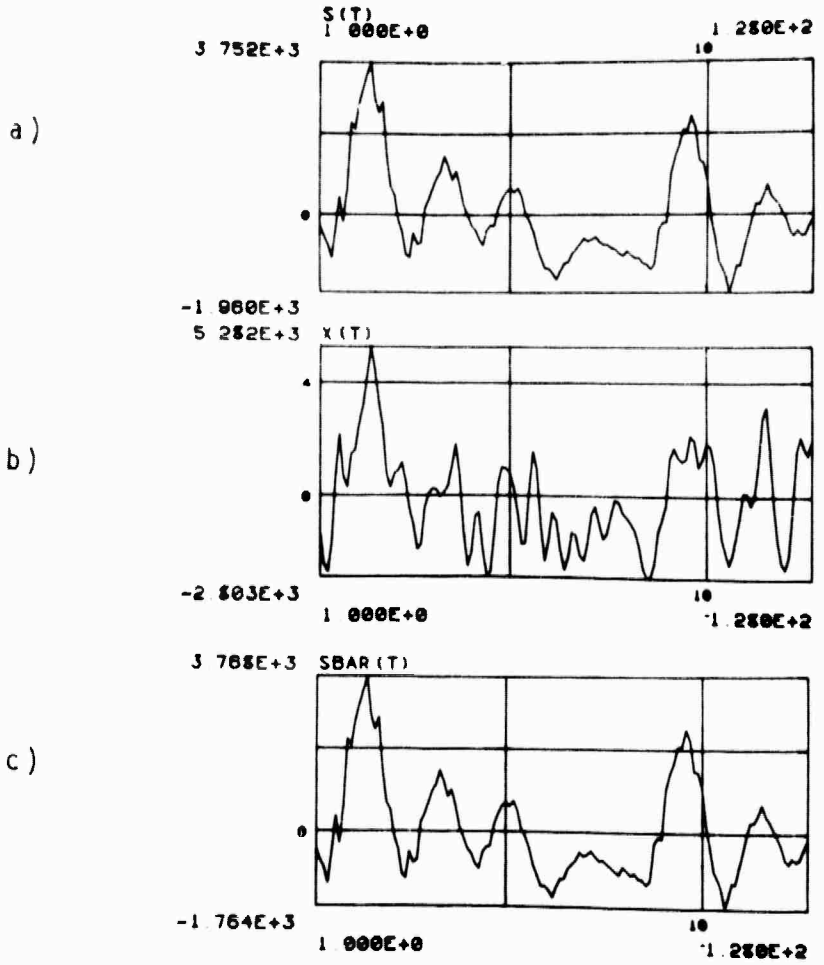


Figure V.5

Signal examples from region where H has converged

a) Original signal b) Noisy signal c) Signal estimate

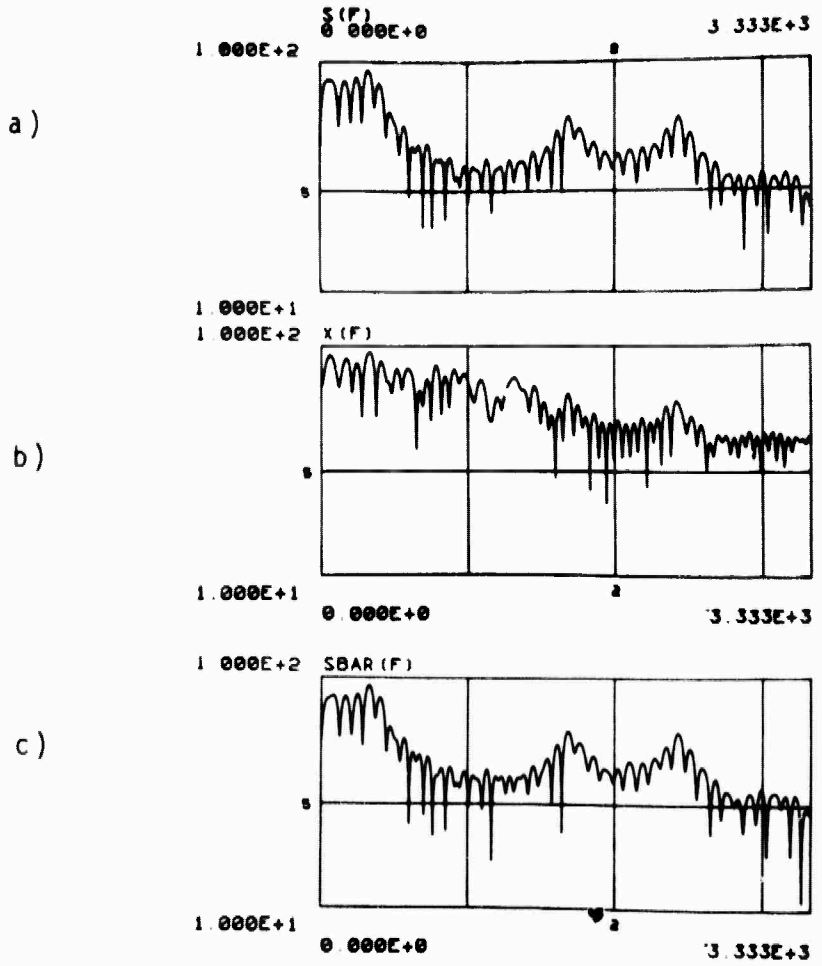


Figure V.6

Spectra of signals in Figure V.5

a) Original signal b) Noisy signal c) Signal estimate

REFERENCES

[1] Removal of Noise From a Voice Signal by Synthesis, N.J. Miller, Ph.D. dissertation, University of Utah, 1973, rep. UTEC-CSc-74-013, May 1973.

[2] Adaptive Noise Cancelling: Principles and Applications Bernard Widrow et.al., Proceedings of the IEEE, vol. 63, pp. 1692-1719, December 1975.

[3] An Adaptive Array for Interference Rejection, Proceedings of the IEEE, vol. 61, pp.748-758, June 1973.

[4] Automatic Equalization for Digital Communication, R.W. Lucky, The Bell System Technical Journal, vol. 44, pp. 547-588, April 1965.

[5] Echo Cancellation in the Telephone Network, Stephen B. Weinstein, IEEE Communications Society Magazine, vol. 15, pp.9-15, January 1977.

[6] A New Digital Echo Canceler for Two-wire Full-Duplex Data Transmission, K.H. Mueller, IEEE Transactions on Communications, vol. COM-24, pp. 956-962, September 1976.

[7] An Adaptive Echo Canceller, M.M. Sondhi, The Bell System Technical Journal, vol. 46, pp. 497-511, March 1967.

[8] Adaptive Noise Cancelling of Sinusoidal Interference, J.R. Glover, Ph.D. dissertation, Stanford University, 1975.

[9] Adaptive Filters, Bernard Widrow, from Aspects of Network and System Theory, edited by R.E. Kalman and N. DeClaris, Holt, Rinehart and Winston, Inc., N.Y.,1970.

[10] Adaptive Estimation With Mutually Correlated Training Samples, T.P. Daniell, Ph.D. dissertation, Stanford University, 1968.

[11] Adaptive Linear Estimation for Stationary M-dependent processes, J.K. Kim and L.D. Davisson, IEEE Transactions on Information Theory, vol. IT-21, pp.23-31, January 1975.

[12] A General Steepest Descent Algorithm, T.M. McSherry, IEEE Transactions on Aerospace and Electronic Systems, vol. AES-12, pp.12-22, January 1976.

[13] On the Design of Gradient Algorithms for Digitally Implemented Filters, R.D. Gitlin, J. Mazo, and M.G.Taylor, IEEE Transactions on Circuit Theory, vol. CT-20, pp.125-136. March 1973.

[14] A Stochastic Approximation Method, H. Robbins and S. Monroe, Annals of Math Statistics, vol. 22, pp. 400-407, 1951.

[15] Stochastic Approximation: A recursive Method for Solving Regression Problems, D.J. Sakrison, in Advances in Communication Theory, vol. 2, Academic Press, New York, 1966.

[16] Stationary and Nonstationary Learning Characteristics of the LMS Adaptive Filter, B. Widrow et.al., Proceedings of the IEEE, vol. 64, pp. 1151-1162, August 1976.

LIST OF FIGURES