Naval Research Laboratory
Washington, DC 20375-5320

NRL/FR/5530--94-9712

Noise Cancellation for CELP Voice Encoders in an F/A-18 Noise Environment

DAVID A. HEIDE

Human-Computer Interaction Laboratory
Information Technology Division

March 18, 1994

Approved for public release; distribution unlimited.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
March 18, 1994

3. REPORT TYPE AND DATES COVERED
INTERIM Jan. 1991 - Jan. 1993

4. TITLE AND SUBTITLE

Noise Cancellation for CELP Voice Encoders in an F/A-18 Noise Environment

5. FUNDING NUMBERS

33904N

6. AUTHOR(S)

David A. Heide

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory
Washington, DC 20375

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/FR/5530--94-9712

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Space and Naval Warfare Systems Command
Arlington, VA 22217

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited.

12. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Because of the severe noise environment in the Navy's F/A-18 jet aircraft, it has always been very difficult to achieve highly intelligible speech using low data rate voice encoders such as the 2.4 kbps LPC-10. As a result, all voice encoding has been done with a high data rate 16.0 kbps CVSD algorithm. The main focus of this research was to develop a technique that could retain the acceptable intelligibility of the high rate encoders while still significantly lowering the data rate required. To achieve these results, a noise cancellation preprocessor was developed to be used in tandem with the new 4.8 kbps CELP encoder that is being implemented in the STU-III.

While many noise reduction techniques were investigated, the best results were accomplished by first reducing the noise through spectral subtraction and then enhancing the important resonant formants of the speech. The results indicated that when the noise cancellation preprocessor was added to the CELP encoder, the DRT intelligibility scores were improved by a significant 5.0 points, making the speech much more acceptable for possible use in the F/A-18.

14. SUBJECT TERMS

Noise Cancellation
Digital Filtering
Spectral Subtraction

15. NUMBER OF PAGES
20

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

CONTENTS

NOISE CANCELLATION FOR CELP VOICE ENCODERS IN AN F/A-18 NOISE ENVIRONMENT

1. INTRODUCTION

Recently, there has been significant interest in 1) reducing the data rate and 2) improving the intelligibility of communication links used by the F/A-18 Navy fighter/attack aircraft. A 16.0 kilobit per second (kbps) continuously variable slope delta (CVSD) encoding scheme is currently used for the voice encoding because the 2.4 kbps LPC-10 (linear predictive coding with 10 coefficients) has been judged to be unacceptable for the F/A-18 environment. The main interest of this report is to investigate a possible compromise between the 2.4 and 16.0 kbps encoding schemes while still maintaining acceptable intelligibility. This compromise will be based on developing a preprocessing algorithm to the Code Excited Linear Predictor (CELP) 4.8 kbps voice encoder now being deployed in the third generation Subscriber Terminal Unit (STU-III) used for secure office-to-office communication. The CELP encoder is described in Federal Standard 1016.

This report discusses:

- principle research objectives,
- problems associated with the F/A-18 noise environment,
- requirements for a candidate noise reduction algorithm,
- performance evaluation issues,
- speech presence detector,
- noise coefficient update rate,
- time domain noise cancellation algorithms,
- frequency domain noise cancellation algorithms,
- results, and
- conclusions.

2. BACKGROUND

2.1 Objectives

As stated above, there are two main objectives in improving the communication links for the F/A-18 aircraft. While it is easy to achieve one objective or the other (either lower the data rate or increase the intelligibility), the focus of this report is to accomplish both. In detail, the two objectives are

- *Lowering the data rate of the voice encoder* — Because of the Navy's dependence on narrowband links, data rate reduction is always a primary goal. By using a 4.8 kbps encoding scheme instead of 16.0 kbps, ten conversations could be supported for every three that are possible now. On the other hand, if the 16.0 kbps channel would still be used for only one 4.8 kbps conversation, the remaining 11.2 kbps could be used for error control coding to make the data much more robust in high bit error rate environments (jamming). Either way, the Navy benefits significantly.

Manuscript approved January 12, 1994.

- *Increasing the intelligibility* — Because of the severe noise environment of the F/A-18, intelligibility improvement is presently a fundamental priority. As measured by the Diagnostic Rhyme Test (DRT), an intelligibility measure described later in the report, the 4.8 kbps CELP has performed remarkably well. In fact, it has scored as high as the 16.0 kbps CVSD in noise-free environments. The problem that this report addresses is whether this performance translates to noisy environments such as the F/A-18 aircraft. As explained below, the intelligibility of voice encoders based on linear prediction such as CELP 4.8 kbps and LPC-10 2.4 kbps typically deteriorates significantly once the noise reaches a certain level.

Previous tests done by Teacher and Coulter [1] show the significant degradation in intelligibility of 2.4 kbps LPC-10 encoded speech at various noisy military platforms. Table 1 gives the DRT scores of unprocessed and processed speech.

Table 1 — Average Unprocessed and Processed (2.4 kbps LPC-10) Speech Intelligibility Scores as a Function of Platform (from Teacher and Coulter [1]).

Platform	Unprocessed Intelligibility	Processed Intelligibility
USS *Saipan*	96	78
USS *MacDonough*	92	74
Jeep	94	78
P-3C Aircraft	96	80
M-60A Tank	87	68
RH-53 Helicopter	89	58
LVTP-7 AMTRACK	72	46

Even with the present day version of the 2.4 kbps LPC-10 algorithm, tests of F-15 jet fighter noise showed a 21 point drop in DRT scores (89 to 68) between unprocessed and processed speech. The goal of this research is to minimize this large drop off in performance that has traditionally hurt LPC-based voice encoders.

2.2 Problems Inherent to the F/A-18 Environment

There are several major factors that account for this substantial decline in intelligibility. Among the most significant are the following:

- *The F/A-18 environment is extremely noisy* — This noise hurts the intelligibility in both analog systems and waveform-based digital encoders, but it is especially harsh for LPC-based voice encoders. Both the 2.4 and 4.8 kbps voice encoders are based on LPC-10. LPC attempts to locate the resonant frequencies of the voice. This location of the resonant frequencies is essential to intelligibility. The analysis, however, becomes very difficult in the presence of noise because the resonant frequencies become hidden by the noise.

Figure 1 shows a typical spectral envelope. The first resonant frequency is easily found, but the second, third, and fourth resonant frequencies are below the spectral floor of the noise. The vowels of words are clear because they are more governed by the first resonant frequency. The consonants, however, depend much more on the weaker upper frequencies. The suppression of the upper resonant frequencies gives the speech a muffled quality so that many words are confused (pick, tick, sick, kick, etc.)

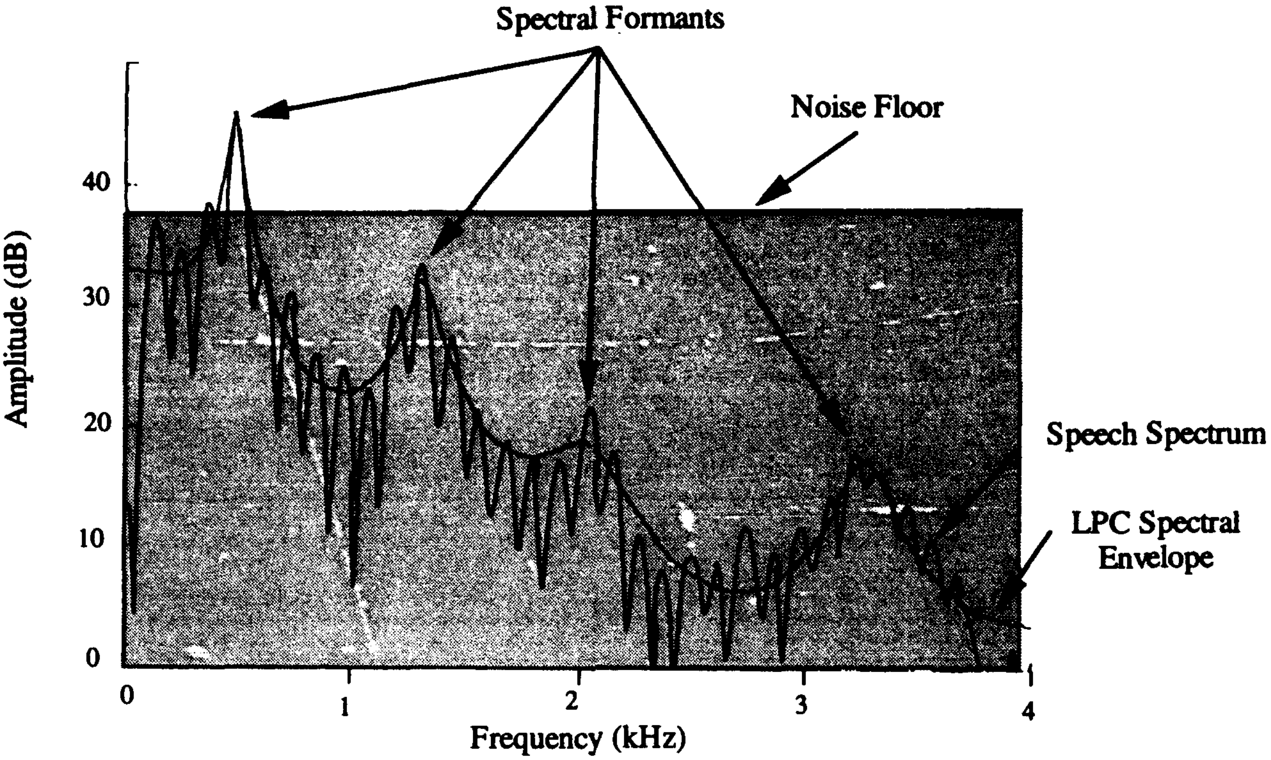


Fig. 1 — Speech Spectrum and LPC spectral envelope with strong noise. (Note that only the first speech formant is above the noise floor; the upper frequency formants are muffled by the noise.)

- *The pilot's oxygen mask introduces distortions into the speech* — Although the oxygen mask helps to reduce the ambient noise of the F/A-18, it also distorts the pilot's speech because of its closed cavity. The speech develops an echo-like quality that reduces the sharpness of the resonant formants, a major problem for LPC analysis. Therefore, even if all of the noise is eliminated by a noise cancellation algorithm, some intelligibility loss would occur.

- *The pilot cannot always speak clearly* — The intelligibility of the voice must be especially good because there will be many situations when the pilot cannot take the time to articulate each word precisely. If the pilot is in the midst of a dogfight or a hard turn, he will not want to have to repeat each word to be understood. Therefore, intelligibility must be maintained under stressful situations.

2.3 Algorithm Requirements

While any candidate algorithm must meet the overall objectives of 1) lowering the data rate and 2) improving the intelligibility of the encoded speech, there are other significant requirements that cannot be ignored. These include the following:

- *No change to the CELP Federal Standard 1016* — The algorithm must not change any aspect of the CELP 4.8 kbps Federal Standard 1016. It must only be a preprocessing algorithm i.e., just feeding the enhanced voice into the CELP encoder. This requirement ensures compatibility with existing communication equipment. This preprocessing requirement also is advantageous because it does not add to the data rate of the voice encoder. It is therefore only limited by computational requirements.

- *Reduce noise without degrading the speech* — Instead of just eliminating as much noise as possible, the algorithm must also consider its effect on the remaining speech. An algorithm could eliminate all the noise, but also make it so that the listener cannot understand any of the resulting muffled words. A compromise must be reached between the elimination of the noise and the effect on the resulting speech intelligibility.

- *Prepare voice for input to the CELP encoder, not the ear* — The goal here goes beyond just noise reduction. The algorithm must also emphasize the speech so that the CELP encoder can perform a good analysis. As stated before, the LPC analysis needs strong resonant frequencies to model the speech well. If the second, third, and fourth resonances are muffled under the noise, the LPC analysis breaks down and jusι models the noise well, not the speech. As a result, all evaluation should take place after the preprocessed speech is fed through the CELP encoder, not just on the preprocessed speech itself. What may sound good to the ear may not "sound" good to the CELP encoder. With this approach, we refine the algorithm to produce good encoded speech, not just good preprocessed speech.

- *Increase intelligibility while maintaining acceptability* — While the main emphasis is on improving the intelligibility of the speech, acceptability (the subjective quality of the speech) is also a major consideration that cannot be ignored. While intelligibility and acceptability usually move in tandem, there are cases where they may not. For example, if the speech is unacceptably harsh, a listener might understand every word but could not possibly listen to it for a prolonged amount of time. The other extreme is a noise removal technique that could remove all annoying aspects of the noise but leave the speech so muffled that nothing is understood. The key is to strike a balance between intelligibility and acceptability when they do not move together.

- *Adapt for changing noise conditions* — The algorithm should recognize changing noise conditions and adapt to them. When speech is absent, the noise estimation can be updated to reflect changing noise characteristics when the F/A-18 is cruising, accelerating, decelerating, etc. Finally, under noise-free conditions when the mission is over and the engine is turned off, the algorithm should default to no preprocessing at all.

- *Use existing equipment* — Any possible improvement should not require additional hardware (second microphone, improved helmet, etc.). As stated above, it should only be a preprocessing algorithm that can be added to the existing software.

2.4 Evaluation

In evaluating the performance of the noise cancellation algorithms, two main criteria are generally used. The first, intelligibility, measures how well listeners can differentiate between one syllable words such as bob, gob, rob, or sip, tip, ship, pip, etc. The standard test for measuring intelligibility is the Diagnostic Rhyme Test (DRT), which consists of three speakers each reading 232 words. The listeners must choose between two possible choices for each word spoken. (Table 2 lists samples of words from a DRT.) The intelligibility score is the percentage of correct words chosen. Enhancing intelligibility is the main focus of this report.

Table 2 — Sample Words of a Diagnostic Rhyme Test (DRT)

gob	-	bob
taunt	-	daunt
moot	-	boot
sheet	-	cheat
jab	-	gab
pot	-	tot
boast	-	ghost
rip	-	lip
said	-	zed
daw	-	gnaw
.		.
.		.
.		.
etc.		etc.

The other main factor to consider is the listener acceptability or overall general quality of the speech. Quality is measured by the Diagnostic Acceptability Measure (DAM) test. This test consists of speakers reading 12 sentences and listeners judging the speech on the basis of 21 different measures relating to the general quality of the speech. Kang and Fransen [2] were able to improve the average DAM score of various military platforms by 6 points through the use of spectral subtraction. While the main focus of this report is to raise the intelligibility DRT scores, the overall quality cannot be ignored.

As stated above, the DRT will be used to measure intelligibility. Unfortunately, no DRT recordings with the F/A-18 noise environment were available. Thankfully, Rome Air Development Center (RADC) provided a DRT from an F-15 fighter jet. The F-15 and the F/A-18 (both manufactured by McDonnell Douglas) have similar acoustics because of their comparably sized twin engines and their matching oxygen helmets (HGU 26/P). Therefore, all evaluation was done using the F-15 as the source of the noise, rather than the F/A-18.

2.5 Speech Presence Detector

In noise cancellation, a very important tool is the speech presence detector. In this report, it is used in two significant places. The first use is to identify noise-only sections so that an accurate analysis of the noise can be done. This way, when the noisy speech is being filtered, only the noise is filtered out. Otherwise, if the noise is estimated when speech is present, speech will also be removed, leading to a loss in intelligibility. The second use of the speech presence detector is involved in the boosting of speech formants. The detector is needed so that no boosting is done when only noise is present.

In this report, the main speech presence detector is based on one from Kang and Fransen [2]. It calculates the energy in the low frequency end of the spectrum (0-1 kHz) and stores it for 128 frames of 180 samples per frame (2.88 seconds). The minimum and maximum energies for the 128 frames are then calculated. If the incoming low frequency energy is above the threshold, the frame is determined to be speech. The threshold is given by

$$\text{threshold} = (\max - \min)/8 + \min.$$

2.6 Noise Coefficient Update Rate

To effectively cancel the F-15 aircraft noise, an effective model of the noise is needed. One of the major considerations in the estimation of the noise model is the update rate of the coefficients to model the noise. One possible approach is to constantly update the coefficients whenever speech is absent. The main advantage of this approach is the very quick response time to changing noise. If the F-15 pilot suddenly accelerates, the noise coefficients are immediately adapted to account for the different noise characteristics. Also, if the pilot lands the plane and shuts off the engines, the noise coefficients automatically revert to zero so that under noise-free conditions, no distortion is introduced by the noise cancellation process.

There is a disadvantage to this approach, however. Since the noise is typically not stationary, the noise coefficients tend to change from frame to frame. This constant adaptation of the noise coefficients causes a "warbling" in the output. While this "warbling" does not affect the intelligibility of the speech, it does affect the acceptability of the speech. Even though the main thrust of this work is to improve the intelligibility of the speech, acceptability cannot be ignored.

Another approach is to initially estimate the noise coefficients and then leave them constant for a specified amount of time until they are updated again. This approach solves the problems of the "warbling" because the noise coefficients are not constantly being updated, but the reaction time to changes in the noise characteristics is greatly reduced. Kang and Fransen [2] developed a solution to this problem by slowly updating the coefficients when only noise was present.

While any final implementation would certainly include this slow updating of the coefficients, it was found in this investigation that only updating the coefficients once at the beginning of each person's DRT list reading was sufficient.

3. TIME DOMAIN (DIGITAL FILTER) APPROACHES

The focus of the report now turns to the main implementation issues. This section presents three approaches aimed at reducing the noise through the use of digital filters. The main objective here is to estimate the filter coefficients that will remove the noise of the F/A-18. The first approach presented derives the filter coefficients by using linear predictive coding (LPC) techniques. The next approach derives the coefficients by using an inverse fast Fourier transform (IFFT) technique. The third approach uses knowledge of the pitch harmonics of speech to derive comb filters to reduce the noise.

After these time domain based approaches are presented, the next section will deal with frequency domain based techniques.

3.1 LPC Residual Approach

In this approach, LPC is used to derive the filter coefficients for the noise cancellation filter. Once the coefficients are computed from noise-only samples, the incoming corrupted speech is filtered with the inverse LPC filter. In typical LPC-based speech compression algorithms, this step is for residual generation. In this approach, however, it is just being used as an efficient way to estimate and remove the spectral envelope of the noise. The filter coefficients can also be easily updated whenever speech is not present.

Figure 2 shows the flow diagram of the noise cancellation process. Figures 3 and 4 show the noise spectrum before and after processing. Notice that the strong spectral components of the noise have been reduced, giving a more flattened spectrum.

The main limitation of this approach is that only the spectral envelope of the noise is modeled because of the limited number of filter coefficients. Even if the number of filter coefficients is increased from 10 all the way up to 40, very limited improvement is achieved.

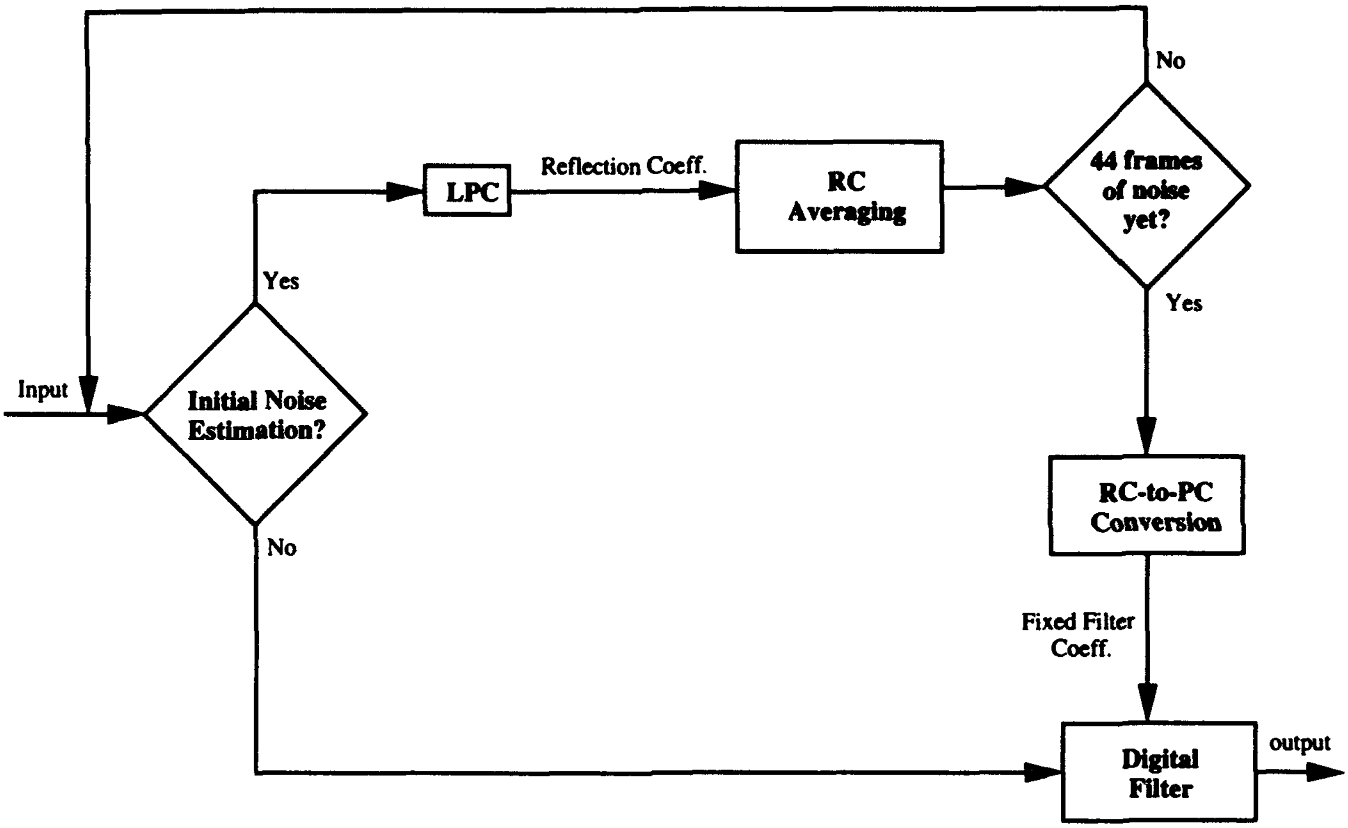


Fig. 2 — Flow diagram for the LPC-based approach to noise cancellation. The initial noise estimation is based on the first 44 frames (1 second) of noise. All subsequent input is then sent straight to the digital filter. Periodically, when speech is absent, the noise estimation process can be repeated to update the filter coefficients.

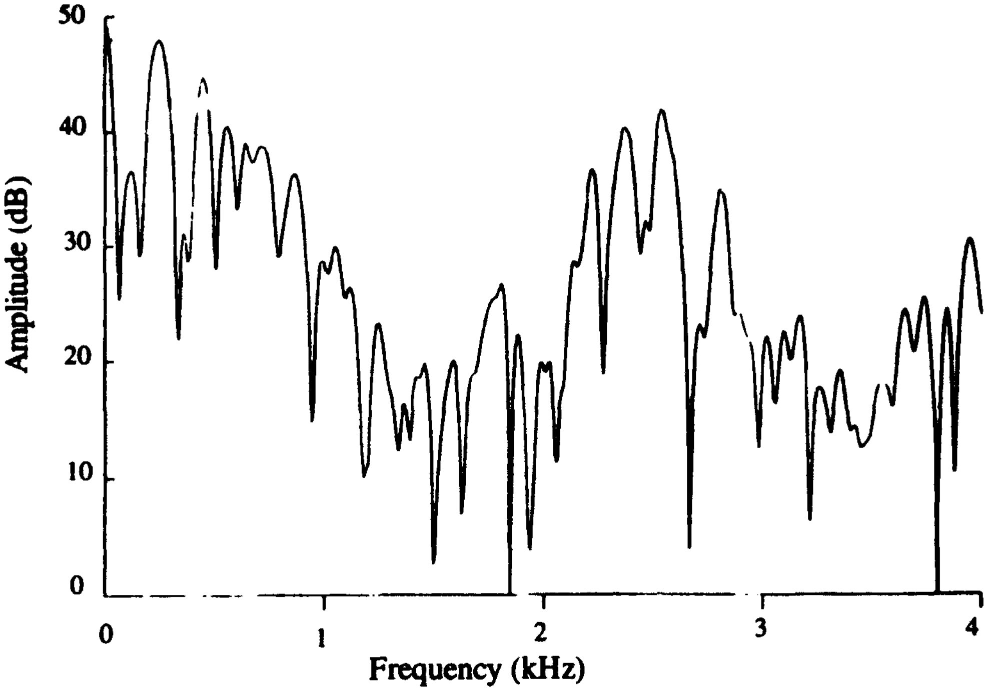


Fig. 3 — Noise spectrum of one frame of noise

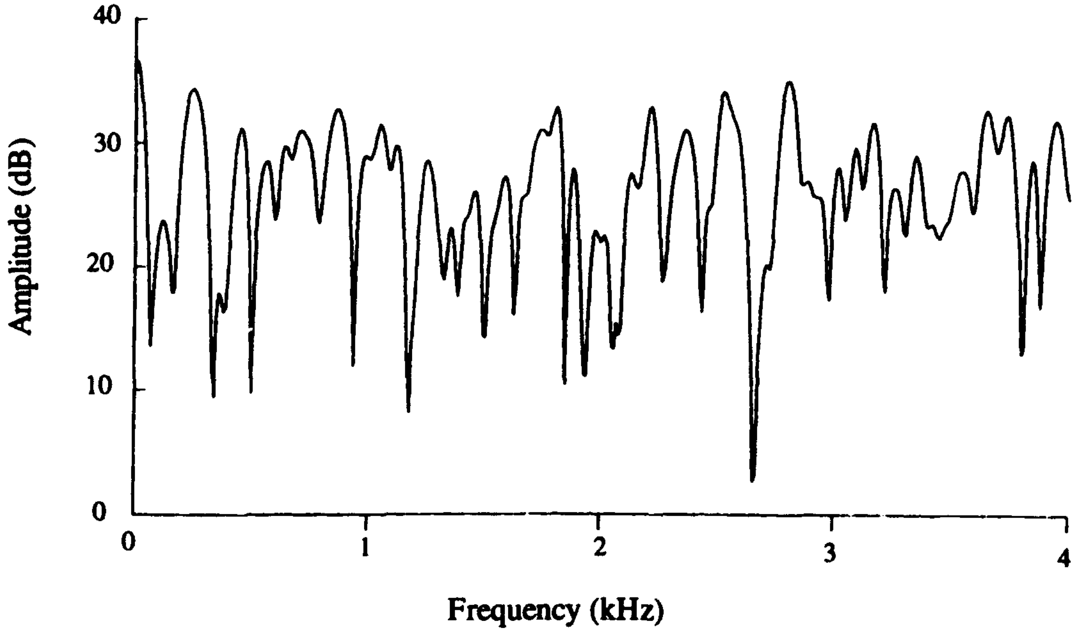


Fig. 4 — Noise spectrum after LPC residual processing. (Note that the two main spectral components of the noise have been reduced to give a more flattened spectrum.)

3.2 Inverse FFT-Based Approach

To achieve a more precise estimation of the noise spectrum, an inverse fast Fourier transform (IFFT) based algorithm was also investigated. This approach first calculates a 256 point fast Fourier transform (FFT), derives the inverse transform 1/H(z), and then computes the IFFT. The result is the impulse response of the noise cancellation filter. Approximately 50 or 60 of the coefficients of the response are saved for generating the output by the standard convolution filter. Even though many coefficients are used, their use is limited to the preprocessing algorithm which does not add to the data rate of the overall system.

Figure 5 shows the filter coefficients (impulse response) derived from this approach. Figure 6 gives the overall flow diagram of the IFFT noise cancellation process.

The main advantage of this approach over the LPC-based approach is that the FFT calculates the actual spectrum of the noise, not just the spectral envelope. The main disadvantage, however, is the significant increase in processing time necessary to do a 50 coefficient convolution compared with a standard LPC analysis and synthesis as in the approach above.

While this approach does use the FFT/IFFT to calculate the filter coefficients, it is still classified as a time domain technique because all actual noise cancellation is done in the time domain with digital filters. In this sense, it is entirely different from the frequency domain method based on spectral subtraction discussed in Section 4.

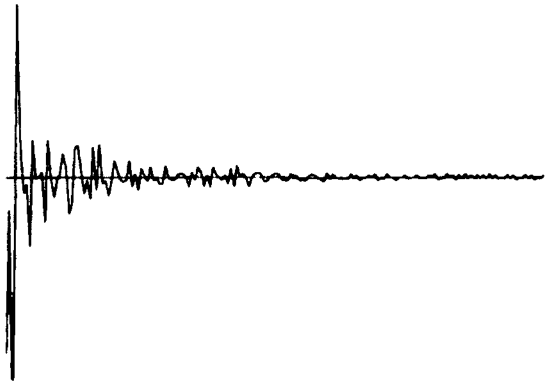


Fig. 5 — Filter coefficients (impulse reponse) derived from the IFFT approach

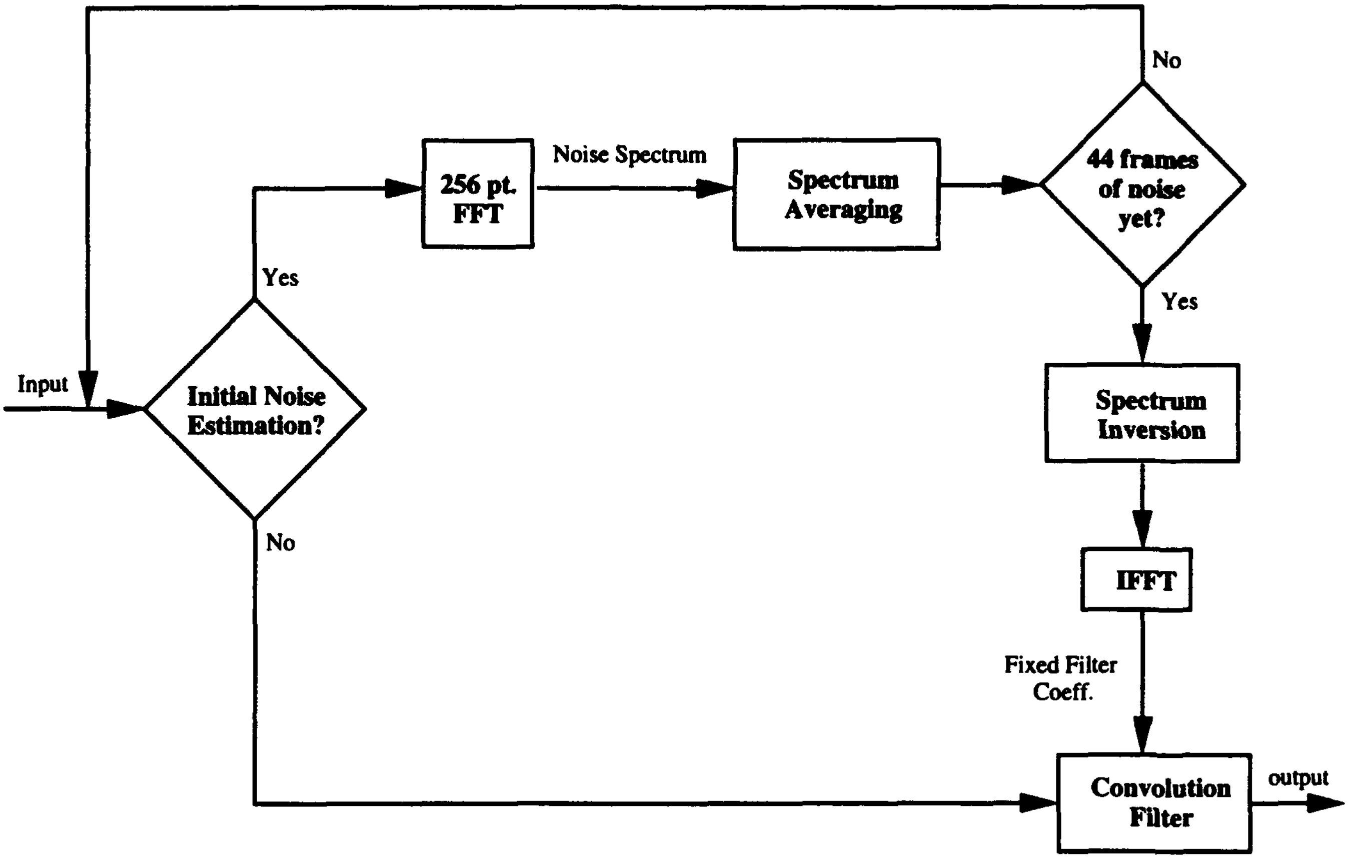


Fig. 6 — Flow diagram for IFFT-based approach to noise cancellation. The initial noise estimation is based on the first 44 frames (1 second) of noise. All subsequent input is then sent straight to the convolution filter. Periodically, when speech is absent, the noise estimation process can be repeated to update the filter coefficients.

3.3 Pitch Filtering Approach

The final time domain filtering approach investigated takes advantage of the pitch harmonic structure of speech. Since the energy of speech is contained in the harmonics of the pitch frequency, any energy outside of these pitch harmonics originates solely from noise that can be filtered out. To filter the noise located outside of the speech pitch harmonics, the noisy speech is first analyzed to determine the frame's pitch period and pitch gain. A comb filter is then used to remove any energy outside of the harmonic bands. Figure 7 shows a typical comb filter derived for speech with a pitch period of 66 samples.

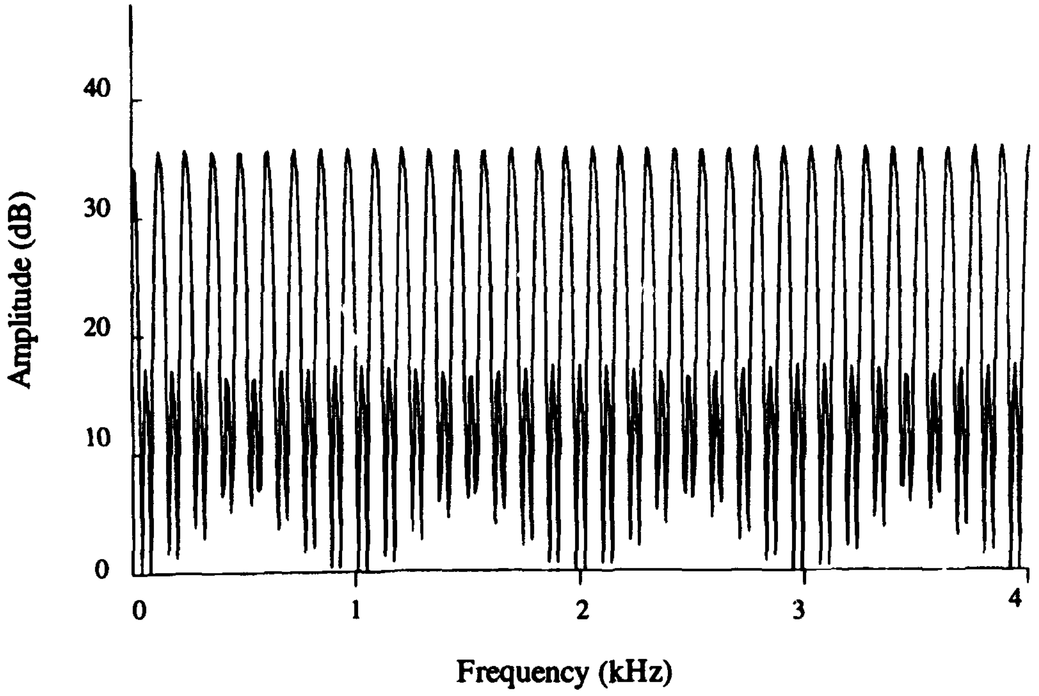


Fig. 7 — Comb filter derived for speech with a pitch period of 66 samples

The main disadvantage to this approach lies in its sensitivity to small errors in determining the pitch period of the speech. Because of the noisy environment of the F/A-18, these small pitch errors occur frequently. If the pitch analyzer is off by two or three samples for the pitch period, the corresponding pitch frequency is also in error. Since the comb filter assumes harmonics at integer multiples of the fundamental frequency, this small error in the original pitch frequency results in a very large error at the higher harmonics. Consequently, the upper bands of the comb filter end up filtering out enough speech that it becomes very unintelligible.

To test whether this approach would succeed even with exact pitch determination, Lim et al. [3] first determined the pitch using noise-free speech. After they found the pitch, they added wideband random noise, which they then attempted to remove. Their results indicated that even when the exact pitch value was known, the filtering process introduced enough distortion that the intelligibility scores still decreased. They attribute this outcome to the fact that speech is not exactly periodic. As a result, this filtering approach was judged to be unsuitable for our applications in the noisy environment of the F/A-18.

4. FREQUENCY DOMAIN (SPECTRAL SUBTRACTION) APPROACH

With limited success using the time domain approaches, the next step was to approach the problem in the frequency domain. In the preceding approaches, the main problem was in estimating the noise cancellation filter coefficients for digital filtering in the time domain. In the frequency domain, however, the main problem is in estimating the frequency spectrum of the noise so that it can be removed through spectral subtraction. Lim and Oppenheim [4] give a good overview of the main techniques involved in spectral subtraction. The main focus of this report is to try to add to these techniques to improve the intelligibility in the F/A-18 noise environment to a more acceptable level.

One of the first issues involved in spectral subtraction is the estimation of the noise spectrum. As previously explained, the best solution for the noise spectrum update rate is a compromise between 1) updating the noise characteristics completely anew every speech-free frame and 2) only updating the characteristics once prior to transmission and then leaving these characteristics constant for a certain time. Because of the relatively stationary nature of the noise in the short DRT list readings used for this investigation, however, the noise characteristics only needed to be updated before each speaker's recording.

4.1 Noise Spectrum Estimation

The noise spectrum was first updated using a discrete Fourier transform (DFT) on a frame of 180 samples calculated at 400 frequency points. The 400 points were chosen to give a 10 Hz resolution over the 4 kHz range. The spectrum (linear response) was averaged over 44 frames (approximately 1 second) of noise before each of the three DRT speakers. The averaging is done over 1 second of noise so that a good overall estimate is obtained. This averaging also helps to eliminate any spurious peaks that may have arisen from only taking the estimate over 1 frame of noise. To further eliminate any spurious peaks after spectral subtraction, a smoothing function was used on the resultant spectrum. However, this resulted in no noticeable difference, probably because the noise spectrum was already smooth enough from 44 frames of averaging.

The initial results from the spectral subtraction were good, much better than anything obtained with the time domain approaches. The DFT, however, made the algorithm run prohibitively slow. By replacing the 400 point DFT with a 256 point FFT, the program run-time was significantly reduced without any noticeable difference in the output speech. Although the 256 point FFT has less frequency precision (31.25 Hz) than the DFT, averaging over 44 frames negated any need for the 10 Hz precision. To implement the 256 point FFT, a trapezoidal window with 76 samples of frame overlap was used [2]. Figure 8 shows the noise spectrum obtained from the FFT.

By using this method, the overall noise level was reduced significantly. However, as with any spectral subtraction method, some amount of degradation of the speech occurred. The main problem is that the peaks of the spectrum of the F-15 noise are also near the typical formants of many speakers. As a result, when the noise spectrum is subtracted, inevitably some of the speech formants are also reduced.

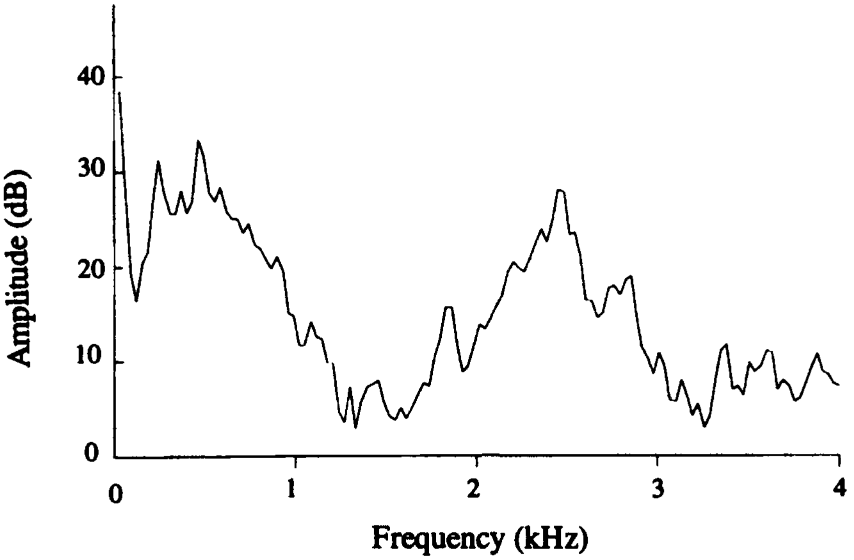
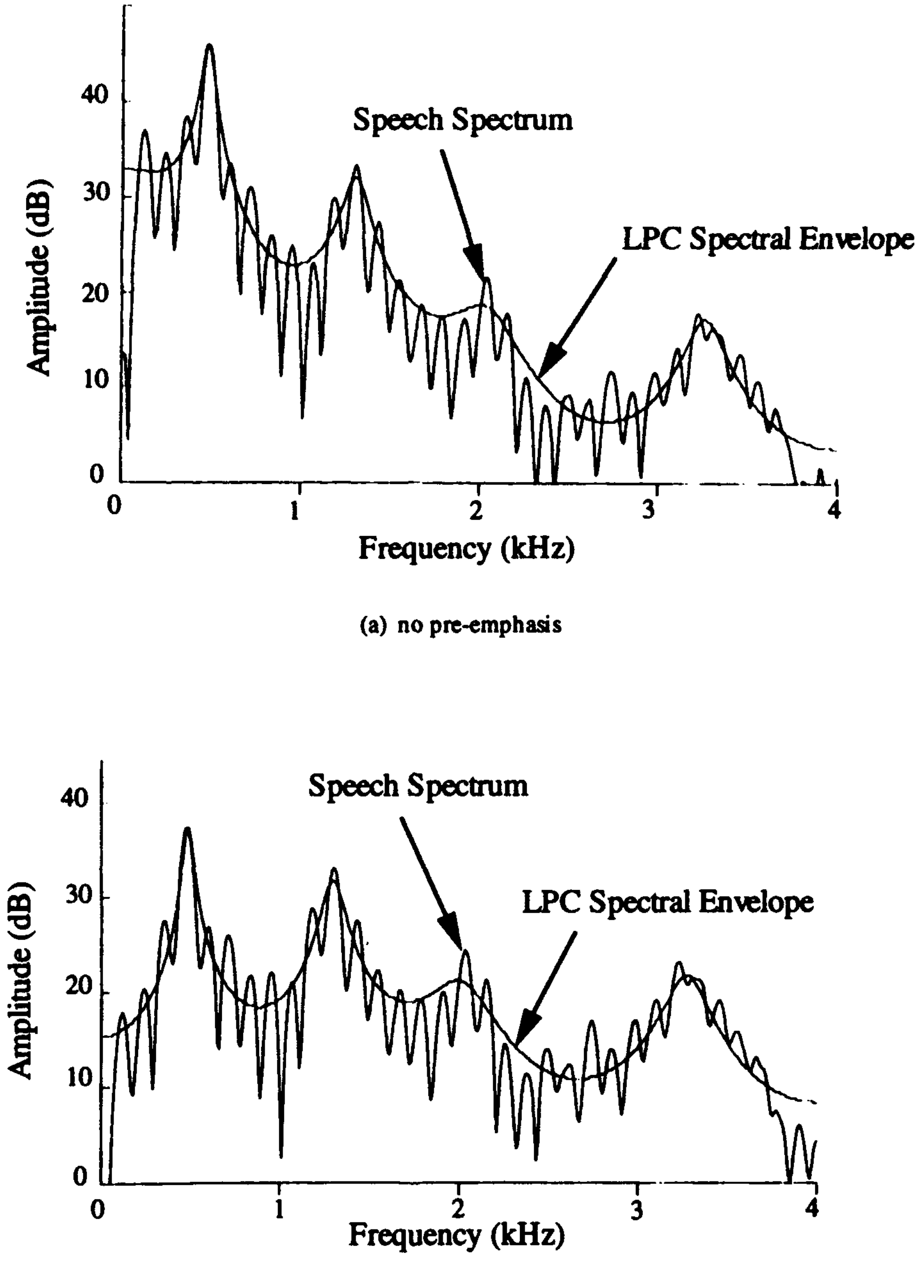


Fig. 8 — Average noise spectrum of 44 frames of 180 samples each, calculated using an overlapped 256 point FFT

4.2 Speech Formant Enhancement

The reduction of the speech formants that results from spectral subtraction is a significant problem for LPC-based speech coders. For this reason, to properly evaluate noise cancellation approaches, the noise-cancelled speech must be first passed through the speech coder before any evaluation takes place. Thus, the noise cancelling process must generate speech suitable for the LPC-based speech coder. Because the LPC process uses an all-pole filter to find the speech formants and corresponding spectral envelope, any reduction in the spectral formants of the speech from the noise cancellation process seriously degrades the performance of the speech coder. To assist in finding the speech formants, typical LPC-based speech encoders almost always use a pre-emphasis filter to boost up the higher frequencies. The pre-emphasis filter is necessary because the second, third, and fourth formants of voiced speech (for even noise-free speech) are usually much lower than the first formant. Pre-emphasis boosts the level of these other formants so the LPC process can detect their location. Figures 9(a) and 9(b) show a typical speech spectrum before and after pre-emphasis.



(a) no pre-emphasis

(b) with pre-emphasis

Fig. 9 — Speech spectrum with LPC spectral envelope before and after pre-emphasis (upper formants boosted)

In noisy environments like the F/A-18, detecting the second, third, and fourth formants becomes much more difficult, even with the pre-emphasis filter. For good intelligibility, it is very important for the LPC process to accurately locate these upper formants. Frequently, however, these formants can be below the noise floor so the LPC process only finds the first formant. In this case, the output speech sounds somewhat muffled, which gives a poor DRT score. The challenge is to find a way to enhance these upper formants so that a brighter sound results, yielding much more intelligible speech.

One approach to boosting the upper formants is to simply use a ramp gain function from 0 dB at 1 kHz to 6 dB at 4 kHz in addition to the pre-emphasis filter of the LPC speech encoder. This gain function, shown in Fig. 10, helps to brighten the speech.

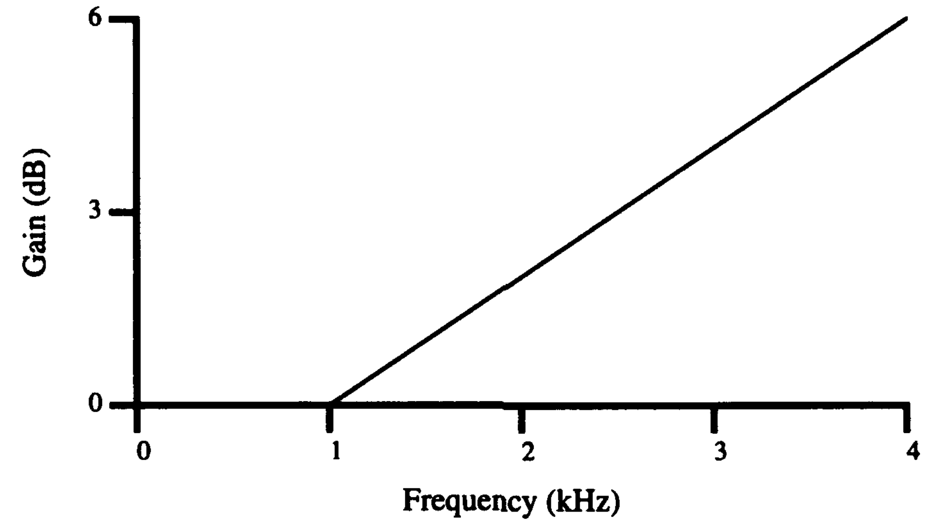


Fig. 10 — Use of ramp gain function to emphasize upper frequencies

Another approach to boosting the upper formants is to search the upper range for the peaks in the spectrum and then boost this small range around these peaks. To find the peaks, the spectral envelope of the spectrum is first constructed by linearly connecting the peaks in the spectrum. The peaks are then easily chosen from the spectral envelope. To boost the upper formants, only the range from 1 kHz to 4 kHz was searched. In this range, the three highest peaks were boosted. Raised peaks of 3, 6, and 9 dB were tested. All three values yielded a much brighter sound, but 6 dB turned out to be the best compromise. The 3 dB value did not yield as much improvement as did 6 dB, but 9 dB tended to be too much of a boost, causing a harsh quality that was irritating to the ear.

Since selective spectral boosting proved successful, the next technique for enhancing the upper formants is to make the *difference* between the spectral peaks and nulls greater. The peaks have already been raised, so the next step is to lower the spectral nulls. As before with the spectral peaks, the spectral envelope is used to find the three lowest spectral nulls from 1 kHz to 4 kHz. This technique has no noticeable difference in the output speech, however, possibly because the real speech spectral nulls were already under the noise floor. As a result, only the spectral boosting was actually implemented. Figure 11 illustrates this technique.

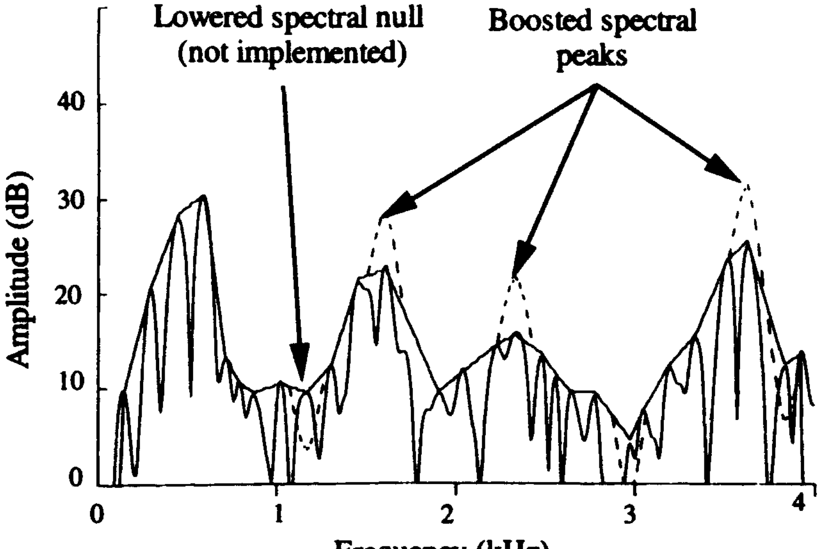


Fig. 11 —Spectral boosting of individual formants

Another consideration in spectral subtraction is the minimum amount of energy left in a spectral coefficient after subtraction. Berouti et. al [5] found that a minor amount of energy should be retained so that the spectral coefficient does not completely go to zero. In this investigation, -30 dB was used as a minimum amount that could result from subtraction. However, no noticeable difference in the output speech was detected.

Figure 12 shows the flow diagram of the final overall noise cancellation algorithm.

5. RESULTS

Although the time domain techniques showed some limited success, informal listening tests undoubtedly indicated that the frequency domain approach reduced the noise much more significantly. Therefore, no formal tests were performed on speech generated by the time domain techniques. All results presented below are from the frequency domain approach.

5.1 Signal-to-Noise Ratio Results

Figures 13(a) and 13(b) show the speech energy of individual words from a Diagnostic Rhyme Test before and after noise cancellation using the frequency domain approach. By using this approach, a gain of approximately 10 dB in the signal-to-noise ratio was achieved.

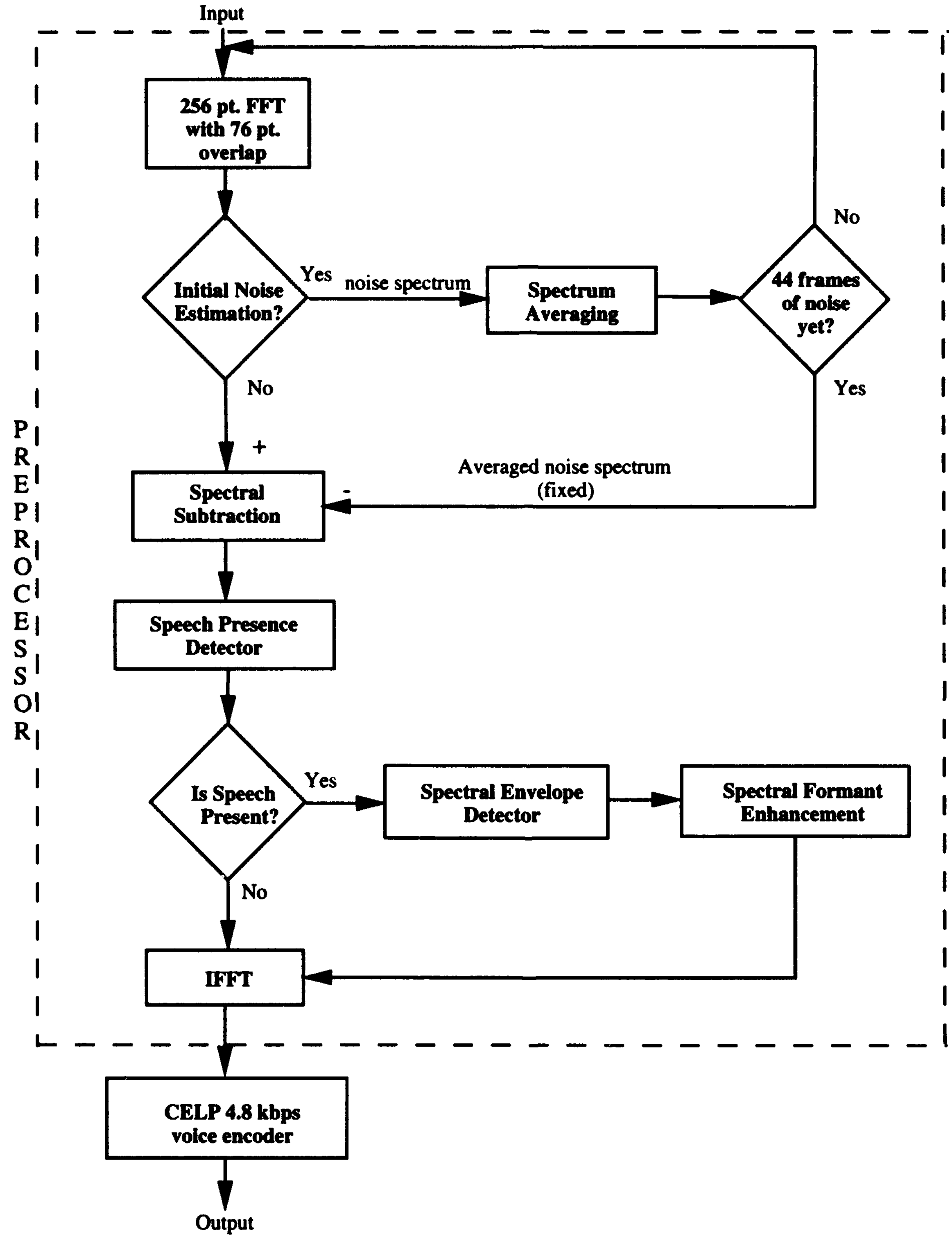
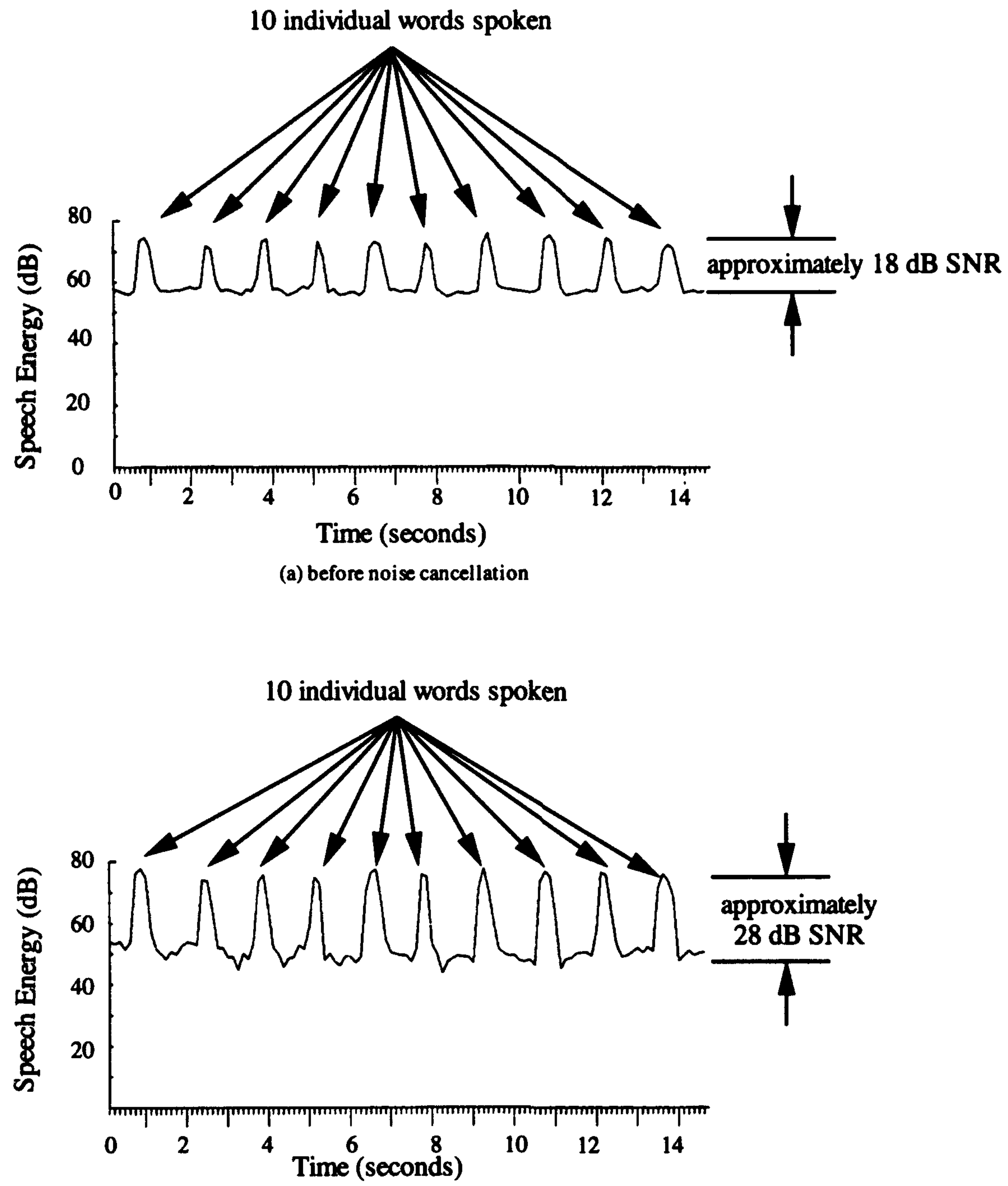

Input
256 pt. FFT with 76 pt. overlap
Initial Noise Estimation?
Yes
noise spectrum
Spectrum Averaging
44 frames of noise yet?
No
Yes
No
+
-
Averaged noise spectrum (fixed)
Spectral Subtraction
Speech Presence Detector
Is Speech Present?
Yes
No
Spectral Envelope Detector
Spectral Formant Enhancement
IFFT
PREPROCESSOR
CELP 4.8 kbps voice encoder
Output

Fig. 12 — Overall flow diagram of frequency domain approach to noise cancellation. Note that the noise spectrum averaging is based on the first 44 frames (1 second) of noise. All subsequent input is then sent directly for spectral subtraction. In addition to its use in the spectral formant enhancement, the speech presence detector could also be used to signal the preprocessor to periodically update the noise spectrum whenever speech was absent.



(a) before noise cancellation

(b) after noise cancellation

Fig. 13 — Speech energy history of 10 spoken words from a DRT list before and after noise cancellation. Note that a gain of approximately 10 dB in the signal-to-noise ratio (SNR) was achieved.

5.2 Intelligibility Score Results

The DRT scores shown in Fig. 14 give the intelligibility of F-15 speech with and without the frequency domain noise cancellation. Note that although the noise cancellation actually hurts the intelligibility of the unprocessed speech, it ultimately improves the intelligibility of the processed speech. This fact relates to the argument that the noise canceller is preparing the speech for the CELP encoder, not the ear. The ear can somehow hear through the noise, but the CELP encoder cannot. The nois canceller prepares the speech for the encoding by subtracting the noise spectrum and then boosting the upper speech formants. The noise cancellation improves the score of the processed speech by a significant 5.0 points.

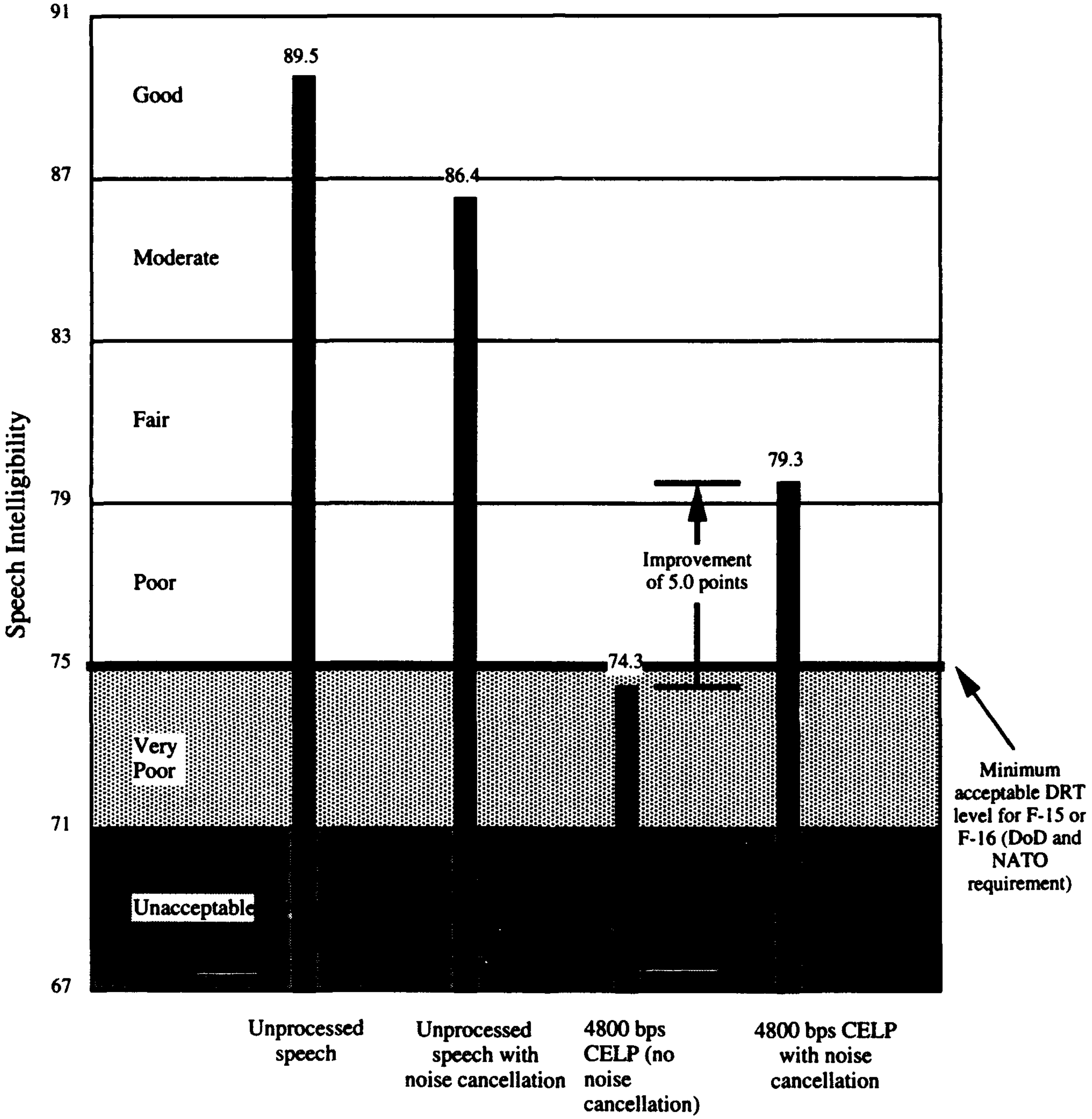


Fig. 14 — Diagnostic Rhyme Test intelligibility scores of F-15 speech

6. CONCLUSIONS

Because of the severe noise environment in the Navy's F/A-18 jet aircraft, it has always been very difficult to achieve highly intelligible speech using low data rate voice encoders such as the 2.4 kbps LPC-10. As a result, all voice encoding has been done with a high data rate 16.0 kbps CVSD algorithm. The main focus of this research was to develop a technique that could retain the acceptable intelligibility of the high rate encoders while still significantly lowering the data rate required. To achieve these results, a noise cancellation preprocessor was developed to be used in tandem with the new 4.8 kbps CELP encoder that is being implemented in the STU-III.

While many noise reduction techniques were investigated, the best results were accomplished by first reducing the noise through spectral subtraction and then enhancing the important resonant formants of the speech. The results indicated that when the noise cancellation preprocessor was added to the CELP encoder, the DRT intelligibility scores were improved by a significant 5.0 points, making the speech much more acceptable for possible use in the F/A-18.

7. ACKNOWLEDGMENTS

The author thanks Sharon James and Tim McChesney of the Space and Naval Warfare Systems Command for support of this research. The author also thanks George Kang of NRL for his many suggestions.

REFERENCES

1. C.F. Teacher and D.C. Coulter, "Performance of LPC Vocoders in a Noisy Environment," IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP-79), Washington, D.C., April 2-4, 1979, pp. 216-219.

2. G.S. Kang and L.J. Fransen, "Quality Improvement of LPC-Processed Noisy Speech by Using Spectral Subtraction," *IEEE Trans. Acoustics, Speech, Sig. Proc.* **ASSP-37**(6), 939-942 (1989).

3. J.S. Lim, A.V. Oppenheim, and L.D. Braida, "Evaluation of an Adaptive Comb Filtering Method for Enhancing Speech Degraded by White Noise Addition," *IEEE Trans. Acoustics, Speech, Sig. Proc.* **ASSP-26**(4), 354-358 (1978).

4. J.S. Lim and A.V. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," *Proc. IEEE* **67**(12), 1586-1604 (1979).

5. M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of Speech Corrupted by Acoustic Noise," IEEE International Conference of Acoustics, Speech, and Signal Processing (ICASSP-79), Washington, D.C., April 2-4, 1979, pp. 208-211.