AD NUMBER
AD870777
LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

Distribution authorized to U.S. Gov't. agencies and their contractors; Critical Technology; 26 FEB 1970. Other requests shall be referred to Air Force Technical Applications Center, Washington, DC 20333. This document contains export-controlled technical data.

AUTHORITY

UFAF ltr, 25 Jan 1972

AFTAC Project No. VELA/T/0701/B/ASD

CONVERGENCE OF
TIME-DOMAIN ADAPTIVE MAXIMUM-LIKELIHOOD FILTERS
FOR STATIONARY DATA

Technical Report No. 3

SEISMIC ARRAY PROCESSING TECHNIQUES

Prepared by

William H. Swindell

Stanley J. Laster, Project Scientist

Frank H. Binder, Program Manager
Area Code 214, 358-5181, Ext. 6521

TEXAS INSTRUMENTS INCORPORATED
Services Group
P.O. Box 5621
Dallas, Texas 75222

Contract No. F33657-70-C-0100
Amount of Contract: $339,052
Beginning 15 July 1969
Ending 14 July 1970

Prepared for

AIR FORCE TECHNICAL APPLICATIONS CENTER
Washington, D.C. 20333

Sponsored by

ADVANCED RESEARCH PROJECTS AGENCY
Nuclear Monitoring Research Office
ARPA Order No. 624
ARPA Program Code No. 9F10

26 February 1970

services group

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC

CONVERGENCE OF
TIME-DOMAIN ADAPTIVE MAXIMUM-LIKELIHOOD FILTERS
FOR STATIONARY DATA

Technical Report No. 3
SEISMIC ARRAY PROCESSING TECHNIQUES

Prepared by

William H. Swindell

Stanley J. Laster, Project Scientist

Frank H. Binder, Program Manager
Area Code 214, 358-5181, Ext. 6521

TEXAS INSTRUMENTS INCORPORATED
Services Group
P.O. Box 5621
Dallas, Texas 75222

Contract No. F33657-70-C-0100
Amount of Contract: $339,052
Beginning 15 July 1969
Ending 14 July 1970

Prepared for

AIR FORCE TECHNICAL APPLICATIONS CENTER
Washington, D.C. 20333

Sponsored by

ADVANCED RESEARCH PROJECTS AGENCY
Nuclear Monitoring Research Office
ARPA Order No. 624
ARPA Program Code No. 9F10

26 February 1970

services group

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (CONT)

LIST OF TABLES

ABSTRACT

The behavior of an adaptively designed time-domain maximum-likelihood multichannel filter during convergence on stationary data was examined. The covariance matrices of a measured seismic short-period pre-whitened noise sample were used to generate 3300 time points of 13 channel stationary Gaussian data having the measured correlation structure. Using these data, 29-point adaptive filters were computed and applied. Their performance was evaluated as a function of time and compared with the performances of the beamsteer filter and the maximum-likelihood filter generated from the measured matrices.

Beginning with a beamsteer weighted initial filter, the filter was adapted for 3272 points. After 1000 adaptions, the adaptive filter was equally effective as the optimum filter in rejecting high-frequency noise. After 3272 adaptions, low-frequency noise rejection by the adaptive filter was much poorer than that of the optimum filter and not appreciably better than that of the beamsteer filter. Wideband noise reduction obtained by the best adaptive filter was about 3.5 db worse than optimum. The loss in performance was probably caused by incomplete convergence. Estimates of the gradient measurement noise are in good agreement with those predicted by theory.

BLANK PAGE

SECTION I
INTRODUCTION AND SUMMARY

An experiment to examine the behavior of an adaptively-designed maximum-likelihood time-domain filter during its period of convergence while operating on stationary seismic-type data is discussed. The experiment's purpose is to determine those values of the adaption algorithm convergence parameter which obtain the best filter in the least time and to determine how near that filter is to the optimum filter. Mean-squared error is used as the basis of error measurement for the adaption algorithm. The quality of the adaptive filter is judged by its performance relative to the true optimum filter both in wideband noise reduction and in the spectrum of the filter output.

Data used in the experiment were synthesized by computer using normal random numbers shaped by a multichannel filter computed from an actual seismic noise covariance matrix. These data are stationary and have the same correlation statistics as the original noise matrix. Besides being stationary, synthetic data are desirable in that the statistics of the data are defined, not just estimated, by the noise matrix. Knowing the actual rather than the estimated statistics permits calculation of the true optimum maximum-likelihood filter, that filter which has the best average performance on the synthetic data in the long run.

The true optimum maximum-likelihood filter achieves an average of about 8.2 db additional wideband noise rejection over the beamsteer (weighted straight sum) filter. This improvement in noise reduction occurs mainly in the 2.2- to 6-Hz and 0.15- to 0.9-Hz frequency bands.

Only about 4 to 5 db additional noise reduction over beamsteering is obtained by the adaptively designed filters and this improvement is reached after approximately 800 adaptions. Regardless of the convergence parameter used, the adaptively designed filter shows little tendency to increase

services group

this improvement up to the possible 8.2-db level attained by the true filter. The adaptive improvement occurs almost exclusively in the power above 2.2 Hz, where the filter performs about as well as the true filter. Below 0.9 Hz, the adaptive filters show no significant improvement over beamsteer filters.

The additional 4 db of noise reduction obtained by the true filter comes from power below 1 Hz. Examination of the filter coefficients indicates that the ability of the true filter to reject noise at low frequencies is obtained through rather large weights. The magnitude of these coefficients is such that it is estimated that at least 26,000 adaptions would be required for the adaptive filter to attain the same size. There is evidence that the improvement of the true filter at low frequencies is a false gain phenomenon where the filter is working on channel gain inequalities rather than on the spatial organization of propagating energy. If this is indeed the case, it is probable that the true filter designed from gain-equalized data would not give as much low frequency noise rejection and that the adaptive filter output would be more similar to the true filter output.

The best performance, in terms of reduction of mean-squared error by the adaptively designed filter reapplied as a fixed filter, was achieved with a convergence parameter K value of 13 percent of K_{max} (K_{max} is the largest value of K for which the mean-squared error algorithm is stable). Values of K around 2 percent of K_{max} adapted much more slowly while values above 50 percent gave little or no wideband improvement over beamsteering and also distorted the spectrum of the output.

SECTION II
THE EXPERIMENT

A. GENERATION OF SYNTHETIC DATA AND CALCULATION OF THE TRUE FILTER

The synthetic data on which all adaptive design runs were made consist of 3300 time points of 13-channel time series data generated by recursively filtering normally distributed random numbers with a 13-channel, 29-point optimum forward prediction filter. Data generated by this method are stationary and have the same correlation statistics as the noise model from which the prediction filter is computed.[1] A special start-up procedure prevented a starting transient. Figure II-1 shows the first 1000 points of each channel of the synthetic data and illustrates the lack of a starting transient.

The noise model matrix was computed from an ensemble of five sets of noise samples digitized with a 0.072-sec sampling period from a 13-element short-period seismic array. The array geometry is shown in Figure II-2. Data in each set were first prewhitened and then the covariance matrix was estimated out ±29 lags using 3300 time points. The five matrices were then averaged to form the noise model for the data generation process. Prewhitening reduced the power of the microseism peak at 0.2 Hz and the strong spectral line components around 2.5, 4.0, and 5.0 Hz. Response of the 37-point prewhitening filter is shown in Figure II-3. The single-channel (channel 1) power spectrum of the synthetic data appears on the power spectra figures shown in Section III.

Using an infinite-velocity signal model, the optimum 29-point minimum-variance unbiased (MVUB) time-domain filter was found from the noise model matrix.[2] Under the assumption of Gaussian statistics, this filter is also the maximum-likelihood filter.[2] The MVUB filter is constrained to pass any infinite velocity signal without distortion, which is another way of saying that its time response to an infinite velocity impulse is also an impulse.
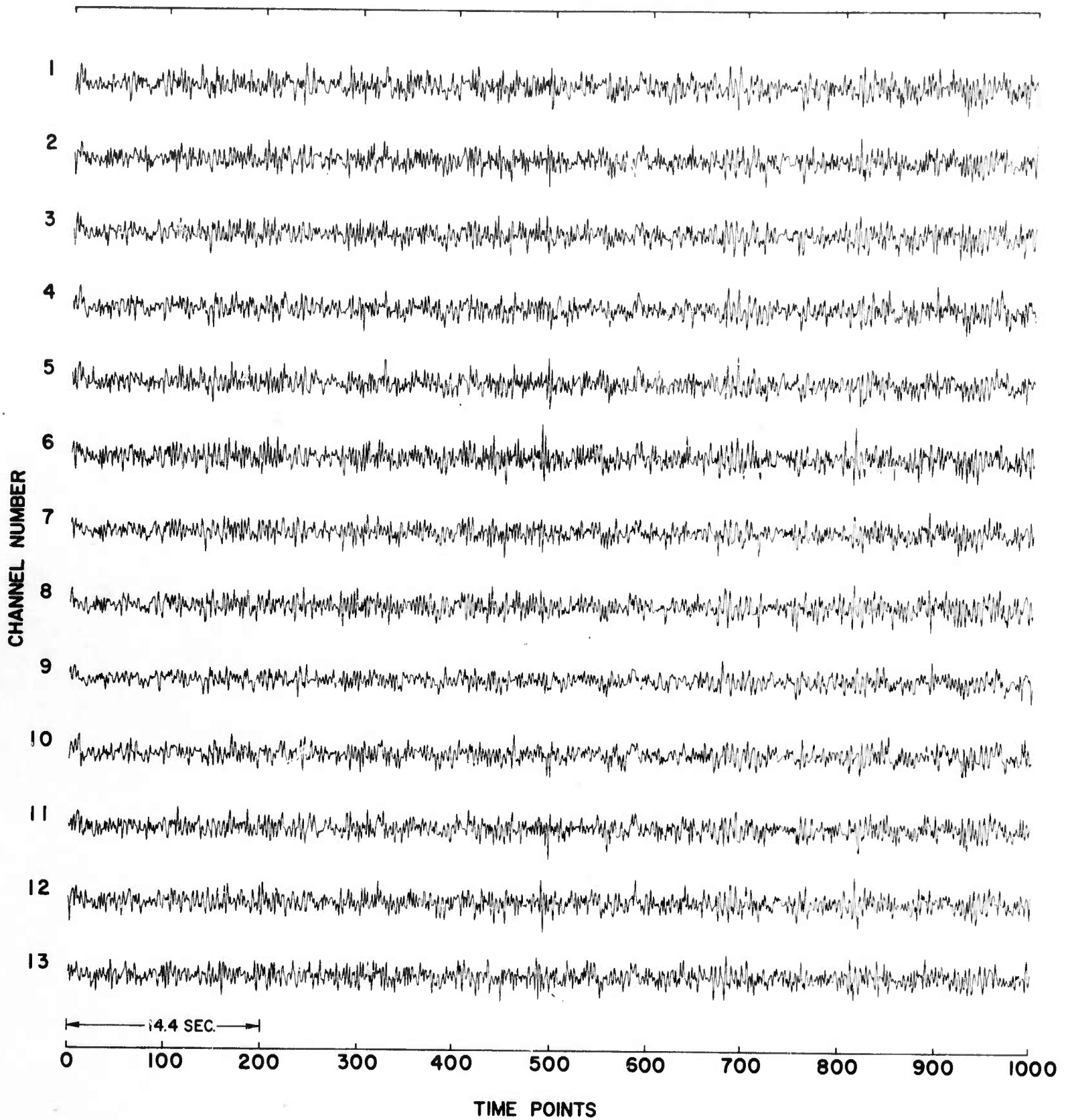
Figure II-1. Synthetic Seismic Data

services group
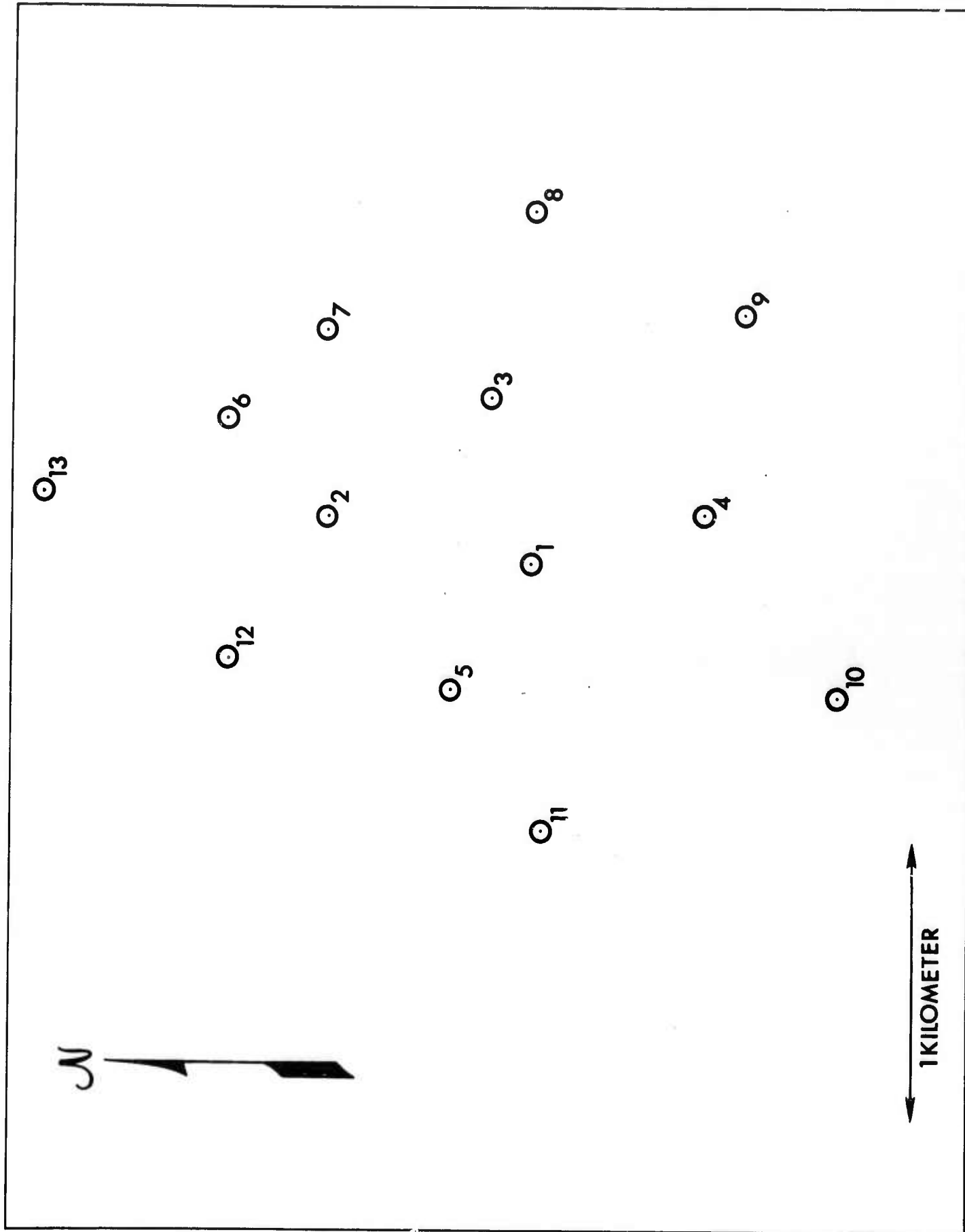
Figure II-2. Geometry of Short-Period Array for Noise Model
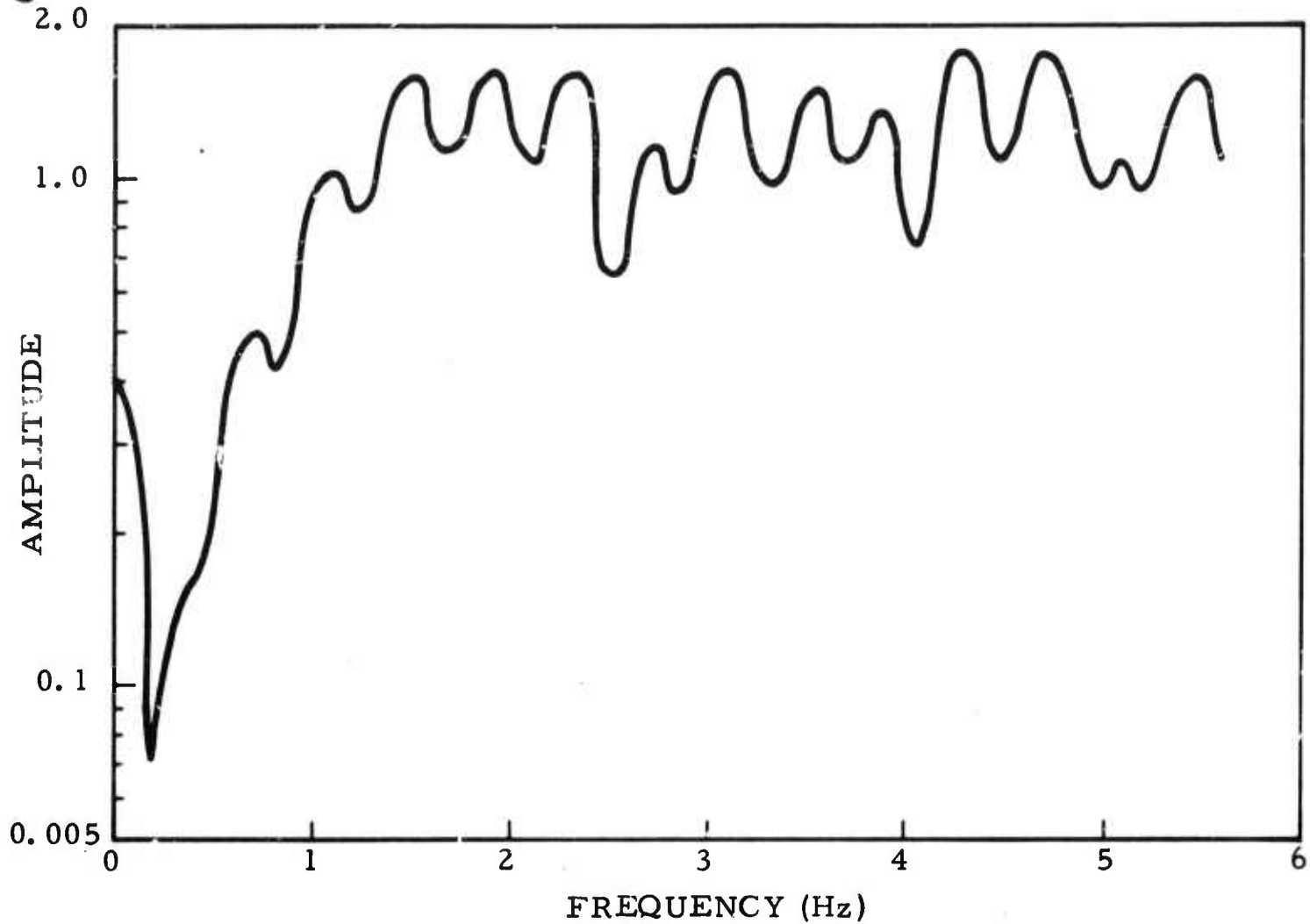
Figure II-3. Frequency Response of Prewhitening Filter

Thus, an N-channel and M-point MVUB filter whose output occurs at point s $(0 \le s \le M-1)$ has an impulse response which is the Kronecker delta:

$$\sum_{i=1}^{N} \sum_{j=0}^{M-1} f_i(j) = \delta_{js} = 1, \qquad j = s$$
$$= 0, \qquad j \ne s$$

   The "beamsteer" filter for an infinite velocity signal is a filter with only one non-zero point whose weight in 1/N for each channel. Its response, letting its output occur at the same point s as the MVUB filter, is

$$\sum_{j=0}^{M-1} h_i = \frac{1}{N} \delta_{js} = \frac{1}{N}, \qquad j = s$$
$$= 0, \qquad j \ne s$$

II-4

For the rest of this report, the MVUB filter designed on the noise model matrix will be referred to as the "true" filter. [3,4]

The true filter and all adaptive filters used 13 channels and were 29 points long (2.088 sec), with the output point centered on the filter at the 15th point (s=15). By designating the time origin to be at t=s, the filters have zero delay. The beamsteer filter then is equal to $1/13\delta_s$ and its output is given by Equation 2-1 below. All adaptive filter design runs, except for two, used the beamsteer filter as the starting filter and, if each filter were not allowed to adapt, its output would be identical to the beamsteer output. The two exceptions were runs in which the true filter was used as a starting filter.

B. ADAPTIVE FILTER ALGORITHM

The adaptively designed maximum-likelihood filters were computed using the "full-gradient" algorithm which has been described in previous reports. [4,5] For convenience, the algorithm is restated here.

The mean of the input data at time t for N channels is

$$\bar{x}(t) = \frac{1}{N} \sum_{i=1}^{N} x_i(t) \tag{2-1}$$

The output of the filter at time point t is

$$y(t) = \bar{x}(t) - \sum_{i=1}^{N} \sum_{j=0}^{M-1} f_i(j) \left[\bar{x}(t-(j-s)) - x_i(t-(j-s)) \right] \tag{2-2}$$

The adaption of the filter after it outputs y(t) to its new state in preparation for computing y(t+1) is, for the mean-squared error (MSE) criterion

$$f_i^{t+1}(j) = f_i^t(j) - 2 K y(t) \left[x_i\left(t - (j - s)\right) - \bar{x}\left(t - (j - s)\right) \right] \tag{2-3}$$

services group

The parameter K is called the "convergence" parameter and is a scalar which controls the relative amount of change in the filter weights in one update step. The MSE algorithm, Equation 2-3, can become unstable when K exceeds some value K_{max}; under this condition the output grows in size exponentially with time[6]. Under the assumption that the data is equalized and using the fact that the trace of the covariance matrix is the sum of its eigenvalues, K_{max} is given by

$$K_{max} = \frac{1}{(N)\,(M)\,\left[R(0)\right]} \qquad (2-4)$$

where R(0) is the average single channel zero-lag autocorrelation of the "input" data. The introduction of the maximum-likelihood constraint into the full-gradient algorithm casts the algorithm into a form of prediction error filtering where the output y(t) is the error in predicting the cross-channel mean $\bar{x}(t)$ using the difference vectors $\bar{x}-x_i$. This is seen in the form of Equation (2-2). The "input" data then is the set of difference vectors $\bar{x}(t)-x_i(t)$ rather than $x_i(t)$ and R(0) is actually $\overline{\left[\bar{x}(t) - x_i(t)\right]^2}$. R(0) was computed by a transformation on the original matrix.

Formulating the algorithm as a prediction error process makes applicable the theory of adaptive predictors. Section III discusses the results and compares some of them with theory.

services group

C. EXPLANATION OF THE PRESENTATION OF RESULTS

Results presented in Section III are shown with three types of figures: improvement plots, power spectra, and filter impulse responses. The output powers of the beamsteer and adaptive filters were averaged over successive blocks of 100 points. The ratio of the average beamsteer power to the average adaptive filter power (called here the "improvement") is plotted every 100 points.[*]

$$\text{IMPROVEMENT} = \frac{\text{AVERAGE BEAMSTEER POWER}}{\text{AVERAGE ADAPTIVE POWER}}$$

Power spectra of a single-channel output and of the beamsteer and adaptive filter outputs were computed to examine the spectral behavior of the filters. Fast Fourier transforms of each 256 time points were taken and the power spectrum for each block was formed. To smooth the spectra, power spectra from four successive 256-point blocks were then averaged. These smoothed spectra were computed for each 1024 points of output to show how the output is affected during convergence. These spectra are shown as the $10 \log_{10}$ (spectral power density); that is, the integral of the area under the curve from zero to folding frequency is equal to the estimated variance of the output.

$$\sum_{i=0}^{127} P(i\Delta f) = \frac{1}{1024} \sum_{j=k}^{k+1023} y^2(j)$$

The impulse responses are plots of the value of the filter coefficients for each channel. The beamsteer filter is not shown but is represented by weights which are all zero except for the weight at the midpoint which has an amplitude of $1/13$ (≈ 0.077).

---

[*] The quantity actually plotted is $10 \log_{10}$(IMPROVEMENT).

In the discussion of the results, the size of the convergence parameter K is given as some percentage of K_{max}. This is done to avoid using numbers which only have significance when compared to the size of K_{max} and to give a feeling of the relative stability of the adaptive process between one run and the next. In some instances the value of K was changed within a run, for instance, from 13 percent to 2.6 percent. For conciseness, this run is designated as (13-2.6).

SECTION III
RESULTS

The efficiency of the adaptive filters is measured by comparing each filter's performance to the performance of the true optimum filter over the same synthetic data. Improvement over beamsteering obtained by applying the true filter to the data is shown in Figure III-1. The average improvement using the ratio of the two powers, each averaged over 3272 points, is about 8.2 db.

Power spectra of the true filter's output are shown in Figures III-2 and III-3. Figure III-2, top, is the spectrum of output points 1 to 1024. Similarly, Figure III-2, middle and bottom, are the spectra of points 1025 to 2048 and 2049 to 3072, respectively. For reference, all power spectra include the spectra of the beamsteer output and of channel 1. Figure III-3 is the average of the spectra in Figure III-2.

The true filter removes substantial noise between 2.2 Hz and 6 Hz. Average power reduction from beamsteering in this band is about 15 db, with the line components at 2.5, 2.7, 3.9, 4.0, 5.0, and 5.5 Hz rejected by 20 to 25 db. Little improvement over beamsteering is evident between 0.9 and 2.2 Hz, with average improvement only about 2 db. Below 0.9 Hz, the true filter again becomes effective and has a reduction of 10 to 15 db between 0.15 Hz and 0.4 Hz, and has 6 to 8 db reduction elsewhere.

Impulse responses of the true filter are shown in Figure III-4. The very large weights may be indicative of the effort required to reject directional power at low frequencies where the array aperture is comparable to the wavelength of the noise. However, rather than true velocity filtering, it is suspected that the improvement from 0.1 to 0.4 Hz are a false gain phenomenon due to interchannel gain inequalities.[7] The evidence for this is that the wavenumber response of the true filter at these frequencies indicates that the system is not doing effective velocity filtering. The implications of

services group

this anomaly will be discussed in Section IV. This low frequency rejection by highly tuned filters, false gain or not, has an adverse effect on the convergence time of an adaptively-designed filter, as will be discussed later in this section.

Various adaptive design runs are listed in Table III-1.

Table III-1

ADAPTIVE DESIGN RUNS

No.	Percent of K_{max}	Notes
1	13	Filter Design
2	53	K_{max} test
3	131	K_{max} test
4	2.6	Filter design
5	13-2.6	Filter design
6	13-53-2.6	Filter design
7	13	True filter adapting
8	2.6	True filter adapting

The first adaptive filter was designed over 3272 time points with the convergence parameter K equal to 13 percent of K_{max}. Its improvement, shown in Figure III-5, increases to a maximum of about 4 db after 800 points and then levels off with little tendency to move closer to the "true" improvement curve, also plotted for comparison. As the adaptive filter converges toward its equilibrium set of weights, its output should produce an improvement curve which would eventually overlay that of the true filter.

Power spectra of the 13 percent K_{max} output are shown in Figure III-6. Figure III-6, top, is the spectrum of the output during the period of most rapid adaption. Below 2.2 Hz, the beamsteer and adaptive spectra almost overlay. Above 2.2 Hz, however, power has been generally reduced 2 to 10 db with the line components having been suppressed by up to 20 db. The spectrum of the second 1024 points shows an additional 1- to 3-db reduction in the high frequencies with essentially little change elsewhere. The third spectrum, Figure III-6, bottom, shows little change from the second, but it appears that noise rejection may be beginning to take place between 0.2 Hz and 1.0 Hz. Filter responses were not computed for the 13 percent run.

The plateau reached by the improvement indicates that the filter obtains some sort of equilibrium and that the time constants associated with any further convergence are much longer than the data at hand. A second explanation is that the convergence parameter is too small and that the apparent plateau is only illusory.

The adaptive filter existing after 3272 updates was applied to the data as a fixed filter in the same way as the true filter. The average improvement using the ratio of beamsteer power averaged over 3272 points to the filtered power averaged over 3272 points was 4.7 db. This is 3.5 db less than the 8.2 db obtained by the true filter.

Two short tests (1200 points) were made to check the value of K_{max} which had been estimated from Equation 2-4. The first run was with K = 53 percent. Its improvement, shown in Figure III-7, fluctuates strongly between positive and negative values, indicating incipient instability. The second test used K = 131 percent. This run begins diverging strongly after 100 points with the output amplitude reaching about a million times the input amplitude at 182 points. Thus, the original estimate of K_{max} appears to be good.

The next test used a K of 2.6 percent. Improvement, shown in Figure III-8, is very similar to the 13 percent run except that the convergence is much slower in the first 1000 points. A plateau is reached after 1500 points, with the average improvement on the plateau about 0.5 db less than for the 13 percent case. Power spectra are shown in Figure III-9. For the first 1024 points, the beamsteer and adaptive spectra are almost identical below 2 Hz. Above 2 Hz the adaptive spectrum is about 1- to 2-db smaller than the beamsteer spectrum, except near the line components where 5- to 15-db improvement is reached. Compared to Figure III-6, top, the 2.6 percent K has not done nearly as well as the 13 percent K. Figures III-9, middle and bottom, show that, while further high-frequency noise reduction is obtained, it is slightly worse than the 13 percent results. Almost no additional noise reduction over beamsteering is obtained below 2 Hz.

Impulse responses of the filters after 3272 adaptions are shown in Figure III-10. Their main feature is that the weights have changed only slightly from the original beamsteer filter. The small weights are very effective against the line spectra, but have no low-frequency capability.

The average improvement of the final 2.6 percent filter when reapplied fixed to the data was 3.6 db. This is 1.1 db worse than the 13 percent filter and 4.6 db worse than the true filter.

The next test used two values of K. Over the first 500 points, K was 13 percent; from point 501 to point 3272, K was 2.6 percent. Improvement of the (13-2.6) run is shown in Figure III-11, and the power spectra are shown in Figure III-12. There is almost no difference in either improvement or power spectra from those of the straight 13 percent run (Figures III-5 and III-6). Most of the output power (about 90 percent) is located below 1.5 Hz; thus, any further improvement would have to come from additional rejection of low-frequency power. The only major difference between the (13-2.6) spectra and the true spectra (Figure III-2) is in the 0-to 0.9-Hz band.

Impulse responses of the (13-2.6) filter after it had undergone 3272 adaptions are shown in Figure III-13. As with the 2.6 percent filters, the filter coefficients have changed only slightly from the starting weights. These small weights give excellent high-frequency rejection, yet it is obvious that they do not have any low-frequency rejection capability.

The average improvement of the final (13-2.6) filter when reapplied fixed to the data was 4.1 db or about 0.6 db less than the straight 13 percent filter and 0.6 db better than the 2.6 percent filter.

Contrary to the appearances of the adaptive filter's improvement plots, it is obvious after comparison to the true filter that the adaptive filters have not reached a true equilibrium. One explanation of this behavior is through the concept of a control system composed of many subsystems, each of which has a transient response controlled by an exponential time constant. These time constants are given[6] by

$$\tau_p = \frac{0.5}{K\lambda_p} \tag{3-1}$$

where λ_p is the pth eigenvalue of the covariance matrix.

The three largest eigenvalues were estimated using the power method and annihilation techniques. They are:

$$\lambda_1 \cong 4.33 \times 10^{10}$$

$$\lambda_2 \cong 4.28 \times 10^{10}$$

$$\lambda_3 \cong 3.49 \times 10^{10}$$

Additional subdominant eigenvalues were not computed because of the rapidly increasing instability of the annihilation method. The transformation matrix used to compute R(0) as mentioned in the previous section is singular. This

reflects the linear dependence introduced by the full gradient algorithm. The difference vector, used to incorporate the MVUB constraint, does not have independent elements. The result of this dependence is that the 29 smallest eigenvalues are zero. These zero eigenvalues imply an infinitely long time constant; however, their processes are virtual in that zero power is associated with them.

The three dominant eigenvalues comprise 29 percent of the trace of the covariance matrix. For a $K = 13$ percent ($K_{max} = 2.4 \times 10^{-12}$), the associated time constants are:

$\tau_1 \cong 37$ seconds (514 time points)

$\tau_2 \cong 37$ seconds (514 time points)

$\tau_3 \cong 46$ seconds (639 time points)

The most likely explanation for the apparent failure of the adaptive filters to converge is the fact the time constants intrinsic to this data are comparable to length of data available.

A more graphic explanation can be given in terms of the magnitude of the numbers involved in the filter update algorithm. Consider the second part of the right-hand side of Equation 2.3. The change in weight in one adaption is

$$\Delta f_i(j) = -2 K y(t) \left[x_i\left(t - (j - s)\right) - \overline{x}\left(t - (j - s)\right) \right]$$
(3-2)

For the synthetic data used here, the following values are typical:

$K_{max} \cong 2.4 \times 10^{-12}$

$R(0) \cong 1.1 \times 10^{9}$

$\overline{|y|} \cong 9.2 \times 10^{3}$

$|x - \overline{x}| \cong \sqrt{R(0)} = 3.3 \times 10^{4}$

The value of $\overline{|y|}$ was taken as the square root of the average power output of the final filters from the 13 percent run when that filter was reapplied fixed. Again chosing a K of 13 percent, plugging these numbers into equation (3-2) gives

$$|\Delta f| \cong 1.9 \times 10^{-4}$$

The largest filter coefficient in Figure III-3 is in channel 6 and is equal to 0.5. Thus, assuming that the adaption was always in the right direction, it would require at least $0.5 \big/ (1.9 \times 10^{-4}) = 2600$ adaptions for the filter to reach its true equilibrium.

The problem then is that the filter weights necessary to suppress low-frequency noise are so large that a great many updates are needed to attain them. Compounding the problem is the strong high-frequency content of the data which, through the difference term in Equation 3-2 above, tends to randomize the direction of adaption, thus slowing convergence. Allowing a 10:1 increase in the number of updates since each one will not be in the right direction, it would take about 26,000 adaptions to reach equilibrium. The only means available of reducing the adaption time is through the parameter K. The remaining filter design runs used generally larger values of K.

The next test attempted to speed up convergence by using a large K even though overall improvement would temporarily suffer. Three different values of K were used: 13 percent (points 1 to 800), 53 percent (points 801 to 2000), and 2.6 percent (points 2001 to 3300). The first K of 13 percent let the filter attain good high-frequency rejection. The 53 percent K was used to force low-frequency convergence, perhaps at the expense of high-frequency rejection. The last K of 2.6 percent allowed settling of the filter and recovery of the high-frequency capability.

Improvement of the (13-53-2.6) run is shown in Figure III-14. During the 53 percent K period, the improvement is severely degraded, reaching -12.9 db at 1400 points, but some recovery is made between 1500 and 2000 points. During the final period while K is 2.6 percent, the improvement becomes about the same as for the 13 percent and the (13-2.6) curves.

The final (13-53-2.6) filter when reapplied to the data as a fixed filter gave an average improvement of 4.6 db which is almost the same as the 13 percent filter.

The (13-53-2.6) power spectra (Figure III-15) have some interesting features. The first spectrum, Figure III-15, top, is approximately the same as Figure III-12, top, except for a 3- to 4-db increase in power below 0.6 Hz. Since this is the spectral region of most power, the overall improvement is drastically affected. The second spectrum, Figure III-15, middle, shows striking changes when compared with Figure III-12, middle. Between 0 to 1.7 Hz, average adaptive power is about 8 db greater than the beamsteer output. Above 3 Hz, the adaptive output retains its high-frequency rejection, although it is 2 to 4 db poorer than the (13-2.6) output.

Figure III-15, bottom, shows the spectrum of the last 1024 points during the 2.6 percent period. Above 3 Hz, the spectrum is essentially the same as the (13-2.6) run. Between 1 Hz and 3 Hz, the (13-53-2.6) output is about 3 db higher than the (13-2.6) output and about 1 to 2 db larger than the beamsteer output. The major difference is in the 0- to 0.4-Hz band where the (13-53-2.6) power has been reduced approximately 4 to 8 db. The improvement curve for the last 1200 points shows little difference from the (13-2.6) output, since the 0- to 0.4-Hz improvement is balanced by the increased power in the 1- to 2-Hz band.

The sudden appearance of significant low-frequency noise suppression seemed to indicate that using a large K did cause more rapid adaption.

This evidence was corroborated by the filter responses (Figure III-16), which were similar to the general shape of the true filters although with much smaller amplitudes. This was particularly true for channels 4, 5, 8, 9, 10, 11, and 12.

The next test demonstrated that the improvement between 0 and 0.4 Hz was probably only a chance occurrence. In an attempt to capitalize on the apparent improvement of the previous filter while using a large K, this run used a K = 53 percent for the first 2500 points and a K = 13 percent for the last 732 points. The improvement is shown in Figure III-17. Improvements generally fluctuate around zero and go strongly negative at 1400 points (-12.5 db), 2400 points (-14.0 db), and at 2500 points (-17.1 db). After K was changed to 13 percent, the improvement probably began stabilizing, but there was not enough data left to provide any evidence.

The (53-13) power spectra are shown in Figure III-18. Within the first 1024 points, the high-frequency noise rejection has become very well developed although it is about the same as for the 13 percent case (Figure III-6, top). The adaptive output power below 2 Hz, however, has increased from 2 to 6 db over the beamsteer power. The spectrum of the second 1024 points shows the same effect, but to a much greater degree. High-frequency noise remains suppressed, but the low-frequency power has increased 5 to 20 db over that for the beamsteer power. The last set of 1024 points has a spectrum which is even poorer than the second set. Adaptive power has increased over the entire spectrum. The line components are still being rejected but by lesser amounts. Interestingly, the adaptive power rises rapidly over the beamsteer power above 5.6 Hz.

The (53-13) filter impulse responses are shown in Figure III-19. Comparing these filters to the (13-53-2.6) filter shows that, other than the appearance of "sawteeth" on the responses, little change has been accomplished by the longer period of rapid adaption. Therefore, the tentative conclusion

that the low-frequency noise reduction may be accounted for by the similarity in shape between the true and the (13-53-2.6) filters must be discarded.

The final (53-13) filter, when reapplied fixed to the data, obtained only 3.1 db average improvement over beamsteering.

From the limited number of tests described above, the values of K which give good results range from about 5 percent to less than 50 percent of K_{max}. The best results are obtained with a K of about 13 percent of K_{max}. Rejection of high frequency power above about 2.3 Hz is always rapidly achieved and this rejection seems rather insensitive to the value of K. On the other hand, rejection of the lower frequencies was much worse than optimum and K values smaller than 50 percent of K_{max} were necessary to prevent performance worse than beamsteering.

There is a second consideration involved in choosing a value for K. K is essentially a gain factor which controls the reaction of the system to an error signal. For a finite memory system as used here, the gradient measurement is made with a finite number of samples and hence contains noise due to random variation. This noise is referred to as "gradient measurement" noise.[6] With a smaller K, system response is more sluggish but excursions away from the optimal filter due to gradient measurement noise are also smaller.

The two final experiments tested the stability of the time filter by using it as the initial filter and allowing it to adopt. Since the filter was already optimal and the data was stationary, the output while adapting gave an estimate of the gradient measurement noise. The two values of K were 13 percent and 2.6 percent of K_{max}. The 13 percent K improvement and output spectra are shown in Figures III-20 and III-21, respectively. While adapting, there is an average decrease of about 0.87 db in improvement with no tendency to degrade further. Comparison of the spectra with those of the true

filter shows that the decrease in improvement came from slightly poorer rejection in the 0- to 1-Hz band. High-frequency rejection is the same.

The filter existing at the end of this run was reapplied to the data as a fixed filter. Average improvement over 3272 points was 7.97 db as compared to the true improvement of 8.22 db, a decrease of 0.25 db.

The 2.6 percent run shows even less change from the true filter. Its improvement while adapting (Figure III-22) deviates an average of only 0.17 db from the true improvement. The power spectra (Figure III-23) are almost an overlay of the fixed true filter spectra.

The final perturbed true filter, when reapplied to the data as a fixed filter, gave an average improvement of 8.26 db, an increase of 0.05 db over the original true filter.

The ratio of the true filtered output power while adapting to the true filter output power gives the relative increase in power due to gradient measurement noise. This ratio, called the "misadjustment" is given by[6]

$$D = \frac{1}{2} \sum_{P=1}^{N-M} \frac{1}{\tau_p} = \frac{1}{2} \sum_{P} 2K\lambda_p \qquad (3-3)$$

which from equation 2-4 becomes

$$D = \frac{K}{K_{max}} \qquad (3-4)$$

For K = 13 percent, D = 0.13 which is the same as 0.53 db. The actual misadjustment of the K = 13 percent run is 0.87 db which agrees well. The K = 2.6 percent run has a predicted misadjustment of 0.11 db and a measured misadjustment of 0.17 db.
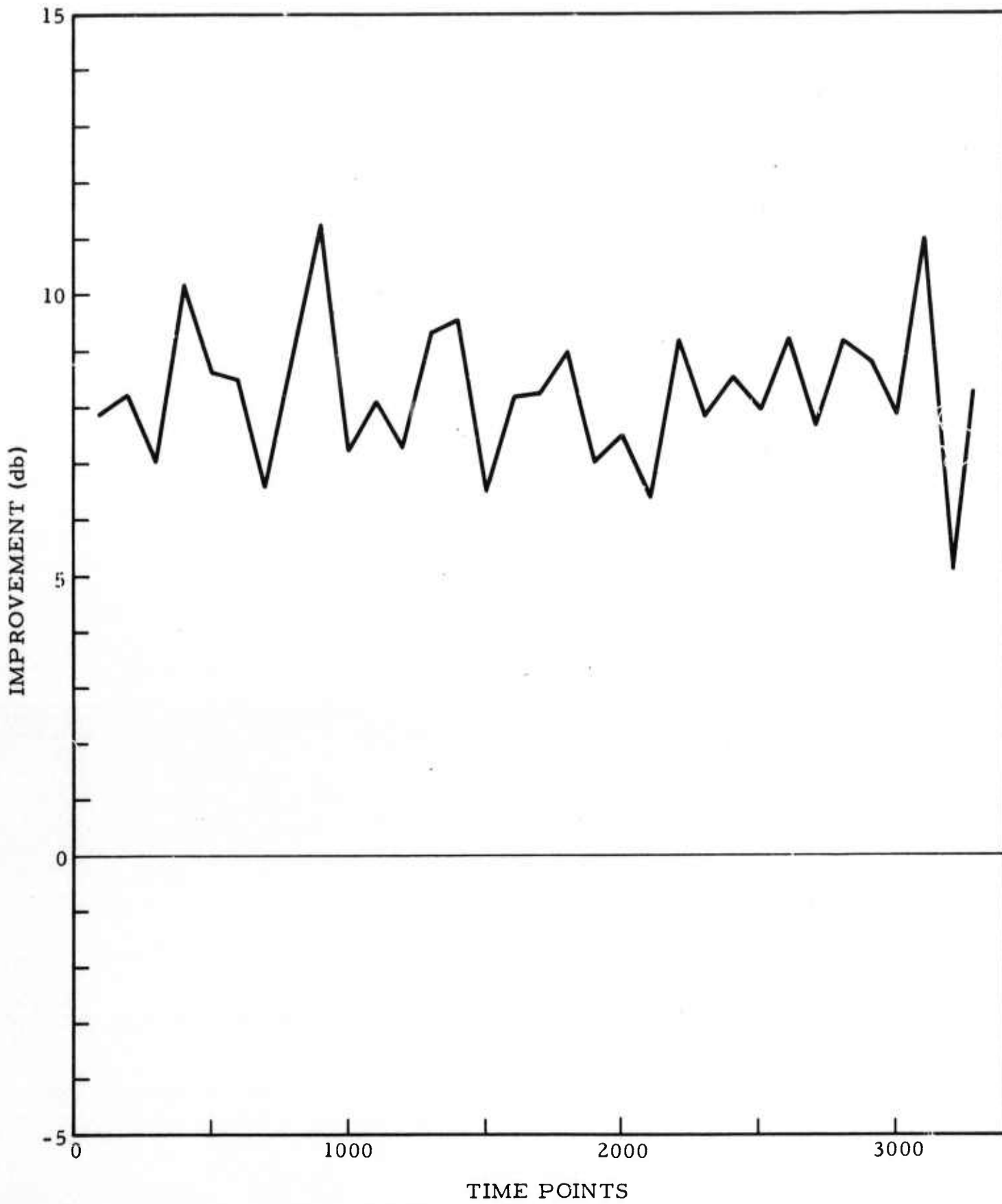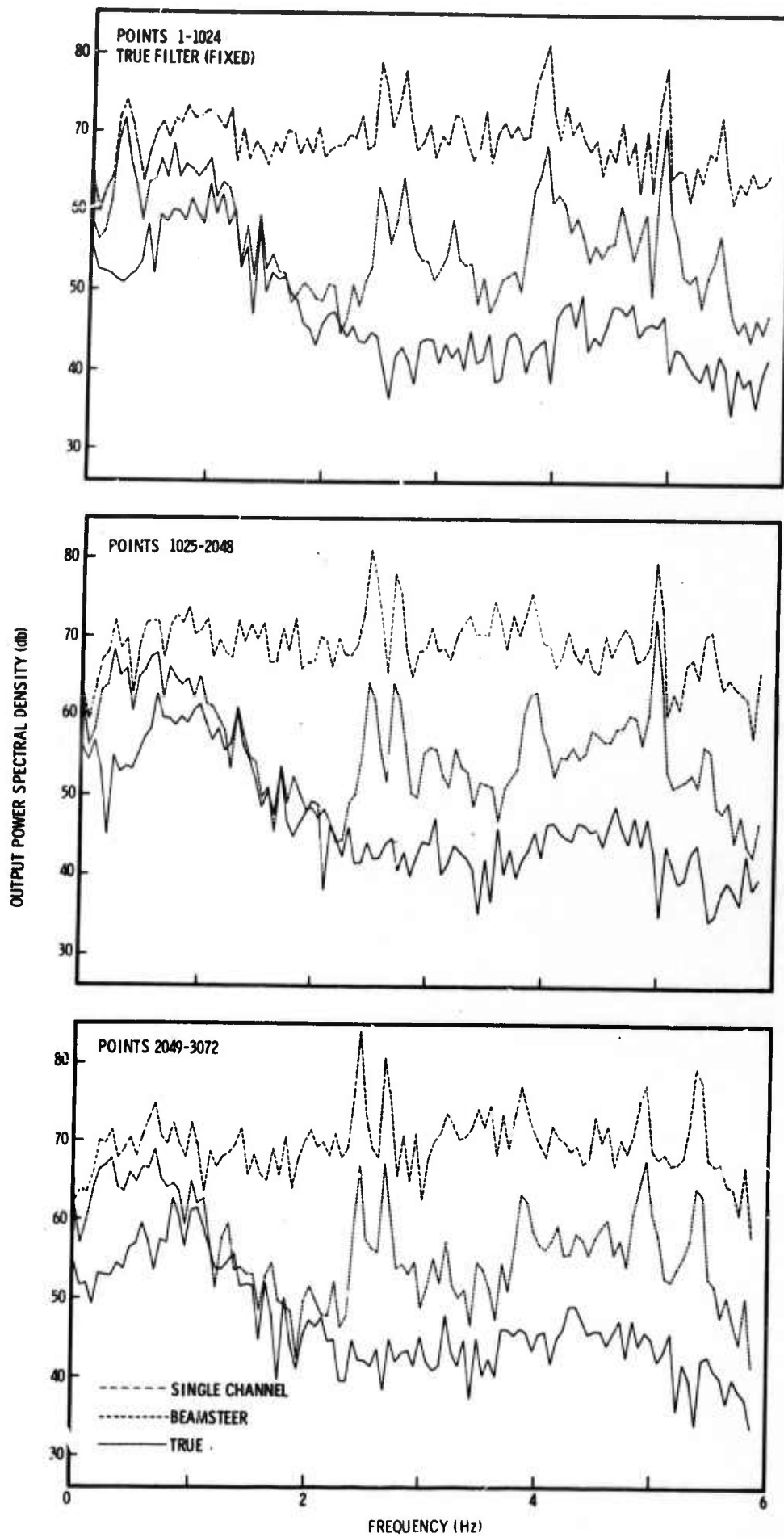
Figure III-1. Improvement of the True Optimum Filter

services group

Figure III-2. Output Power Spectra for the True Optimum Filter
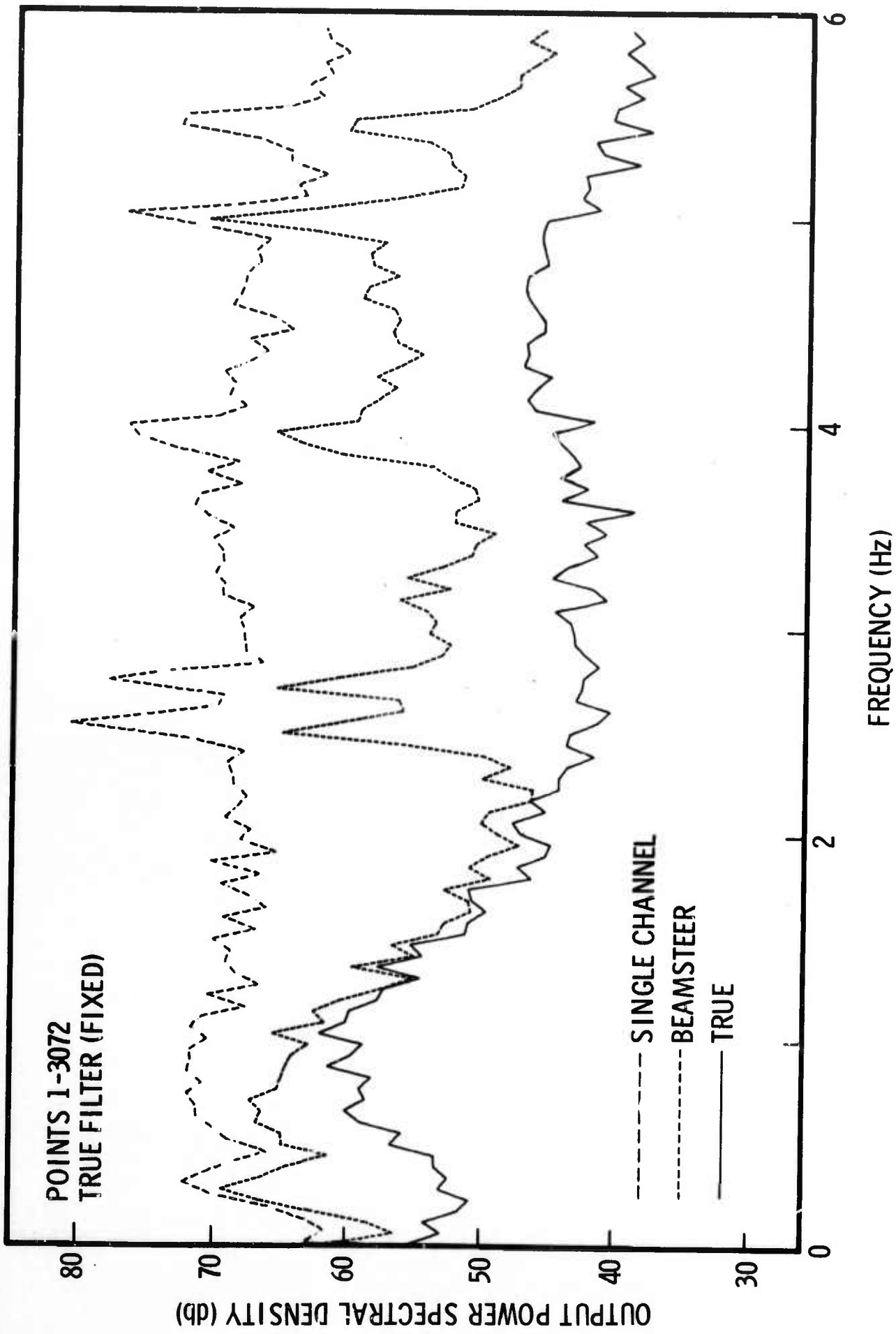
services group

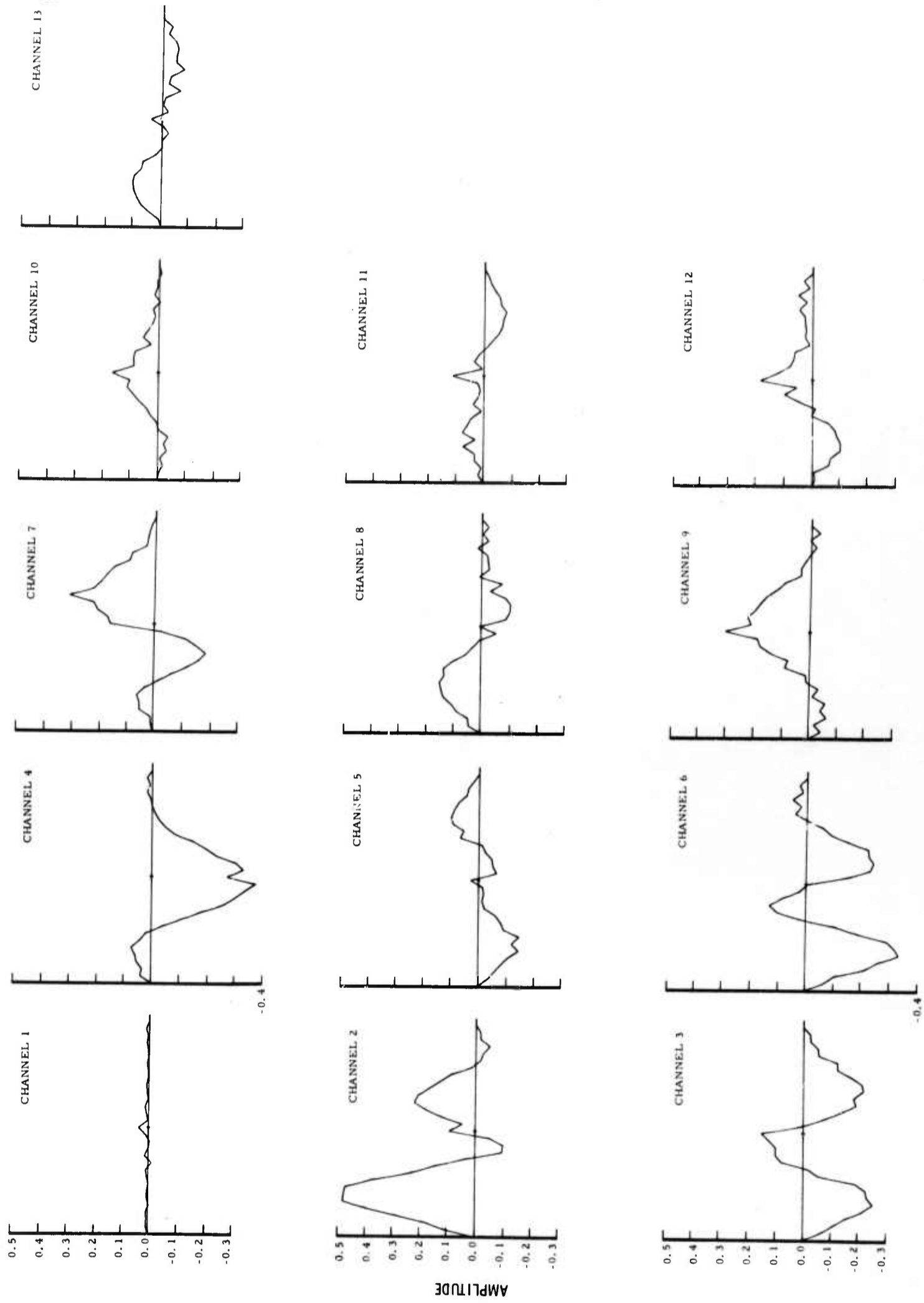Figure III-3. Output Power Spectrum of the True Optimum Filter Averaged Over All Data

services group

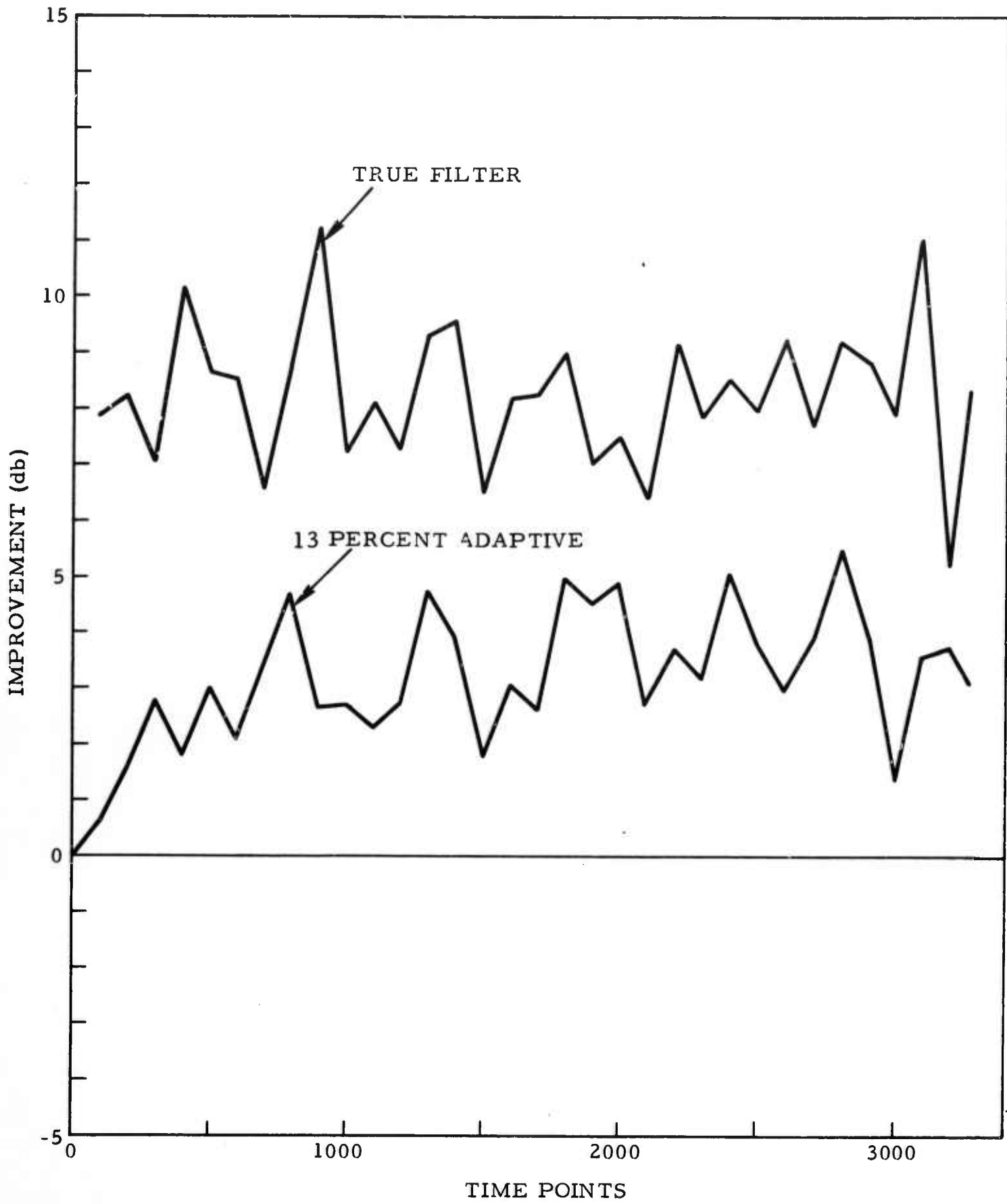Figure III-4. Impulse Responses of the True Optimum Filter

services group

Figure III-5. Improvement of the 13-Percent Adaptive Filter

services group
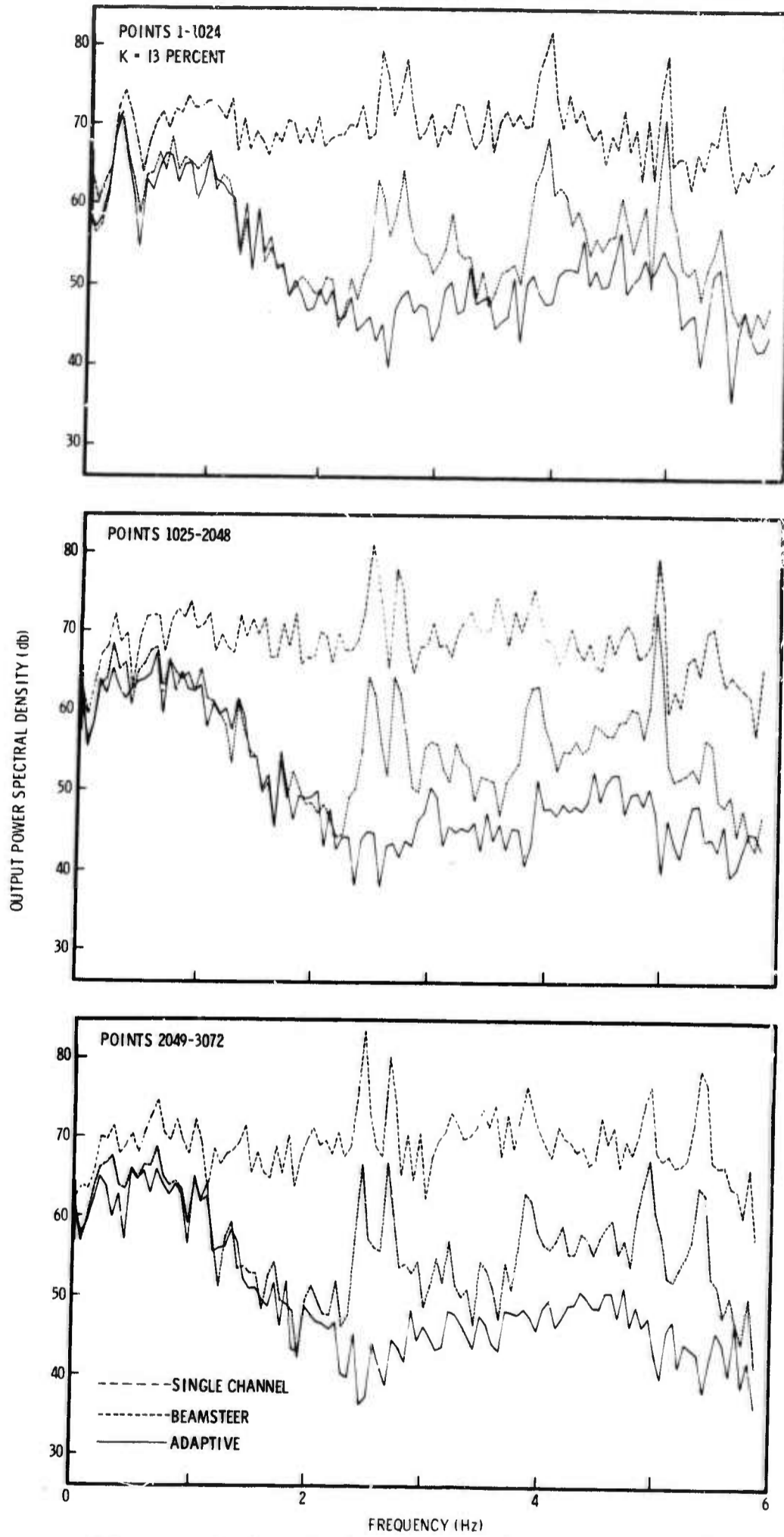
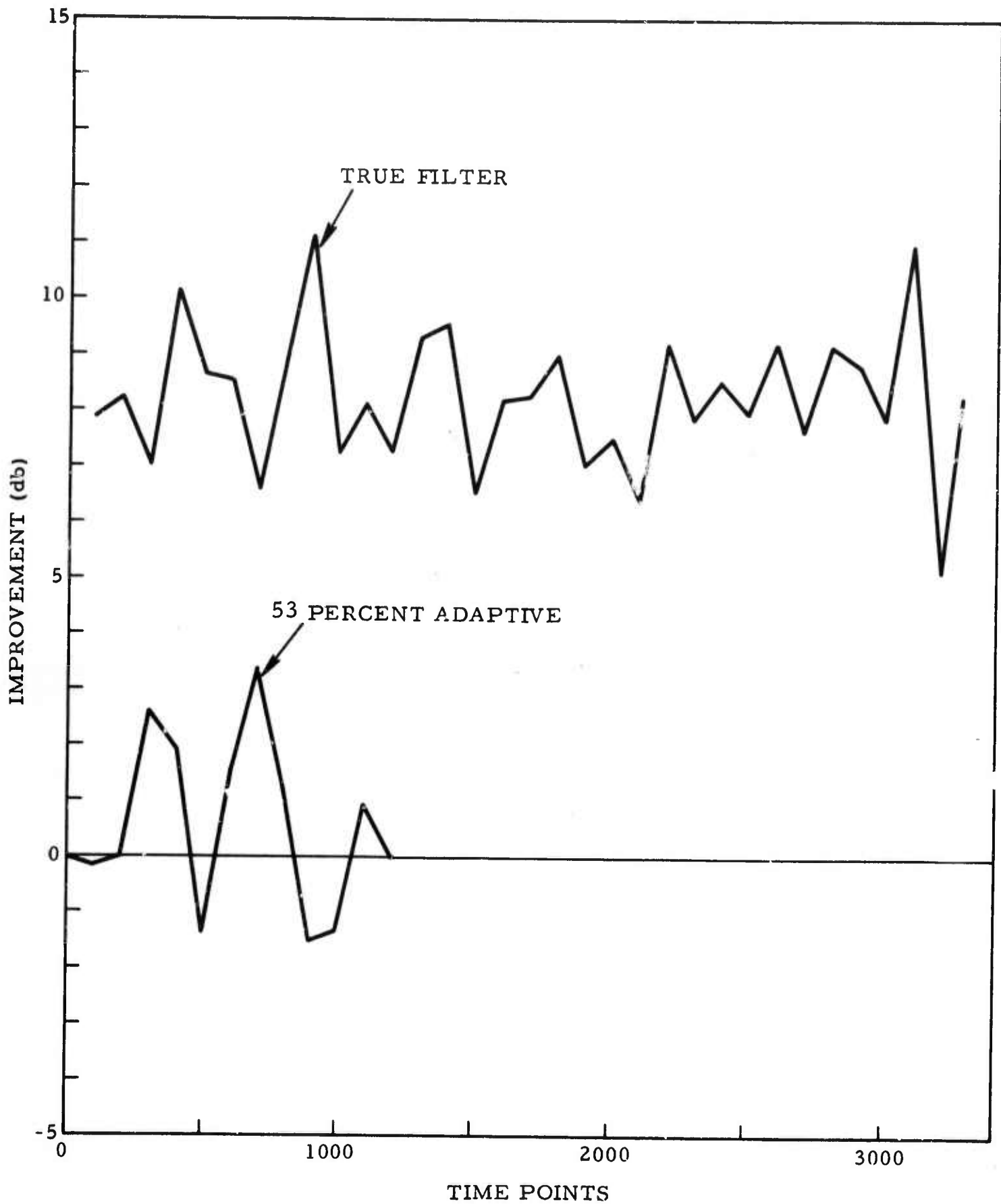Figure III-6. Output Power Spectra for the 13-Percent Adaptive Filter

services group

Figure III-7. Improvement of the 53-Percent Adaptive
Filter During K_{max} Test
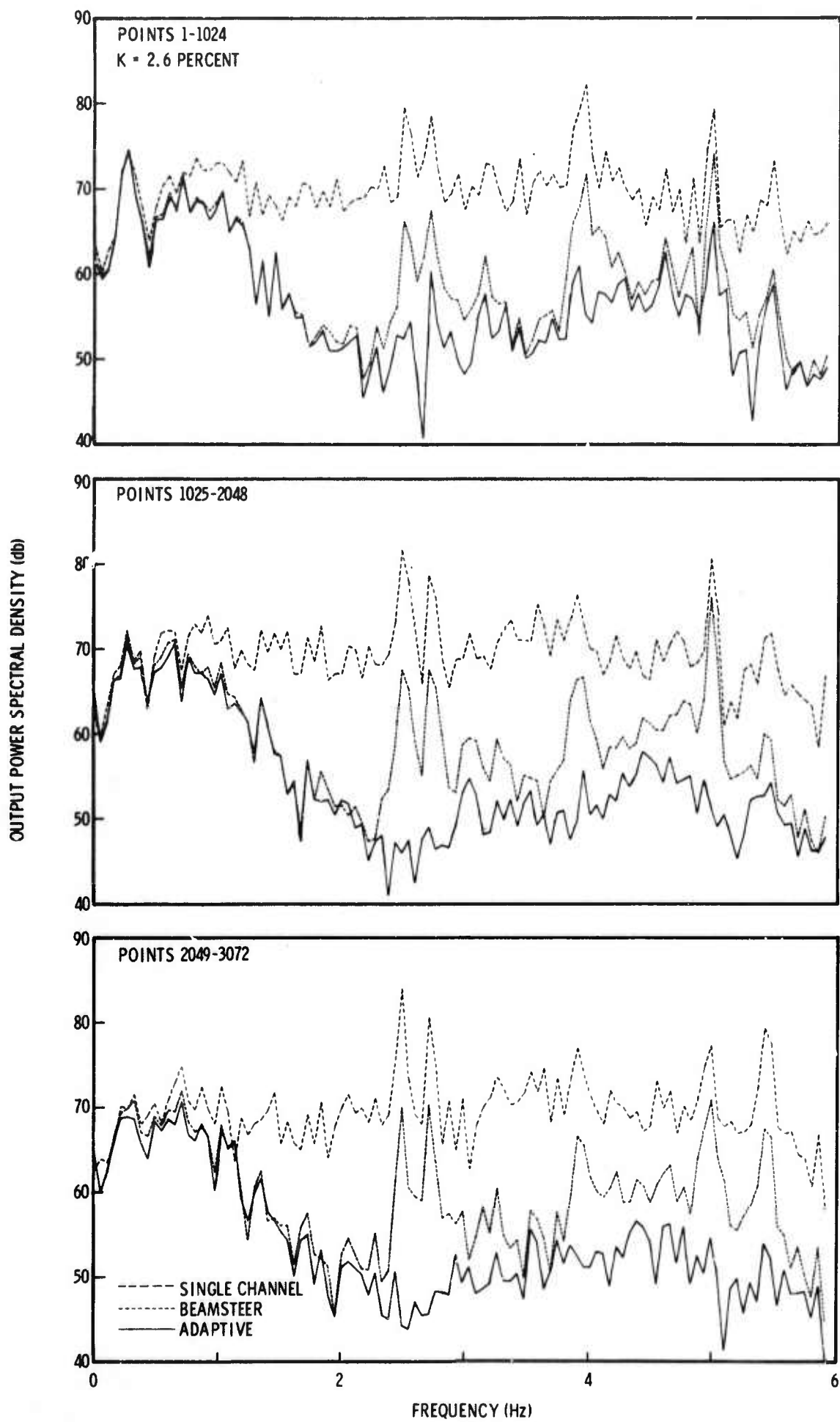
Figure III-8. Improvement of the 2.6 Percent Adaptive Filter

Figure III-9.   Output Power Spectra for the
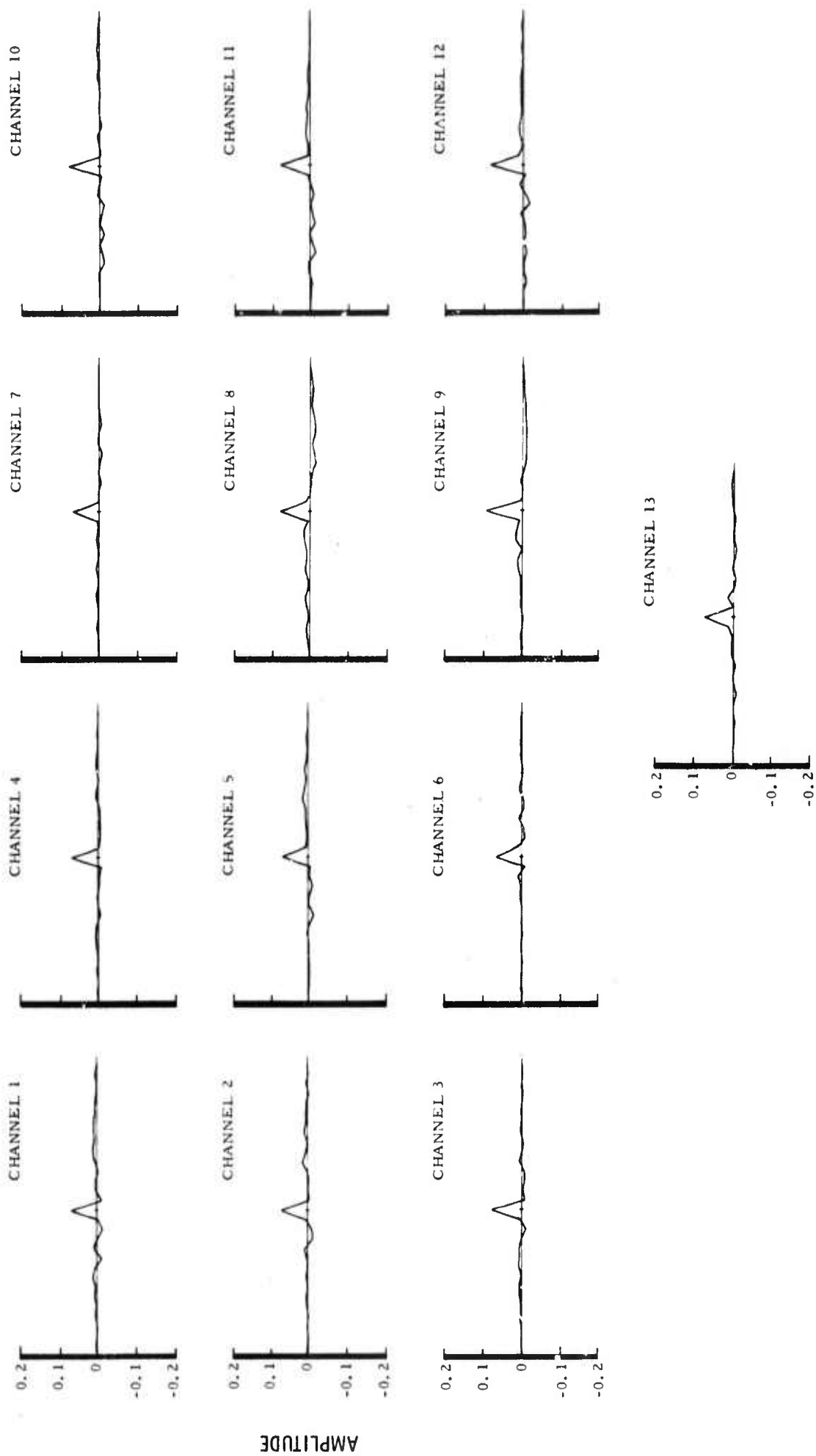2.6 Percent Adaptive Filter

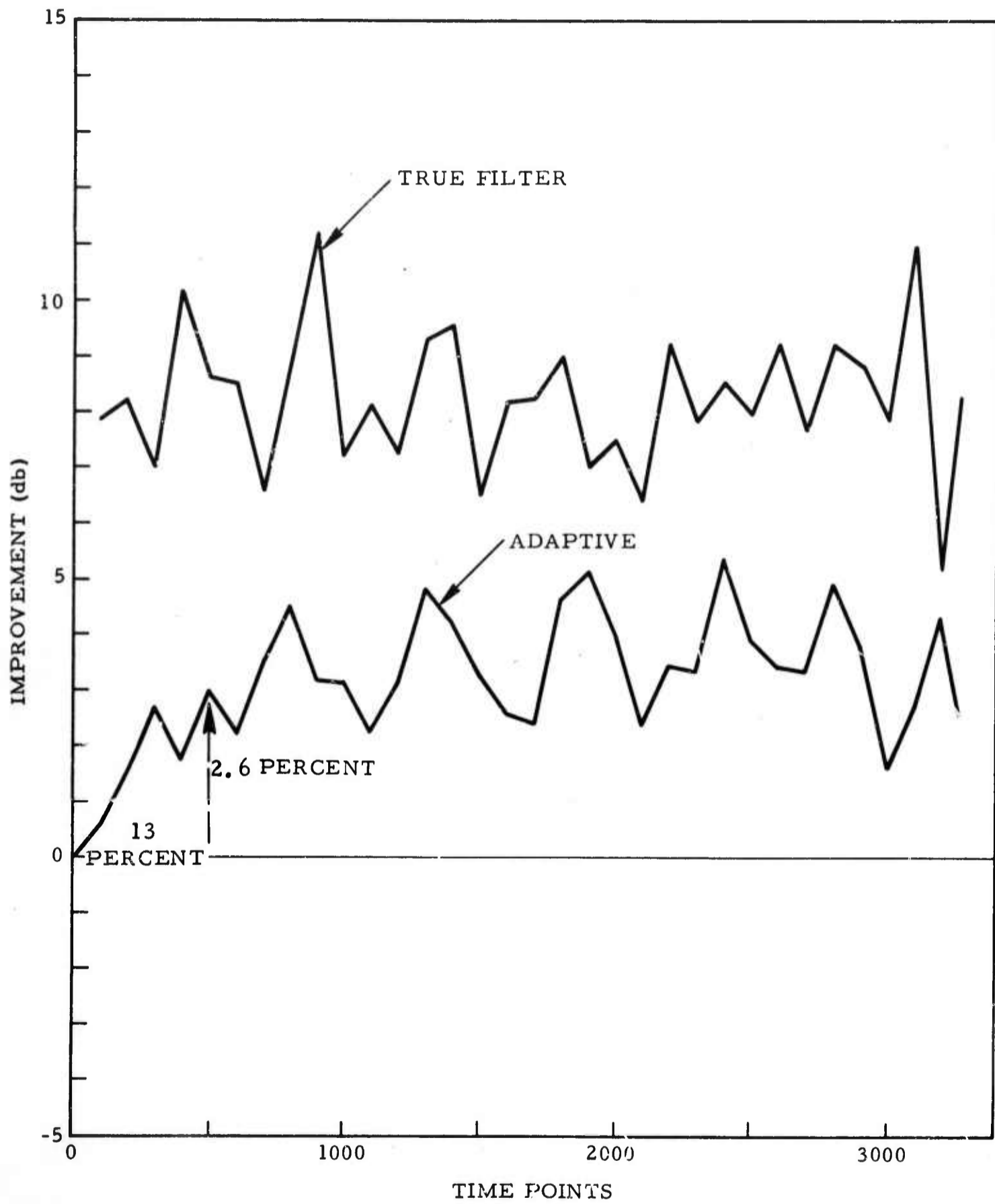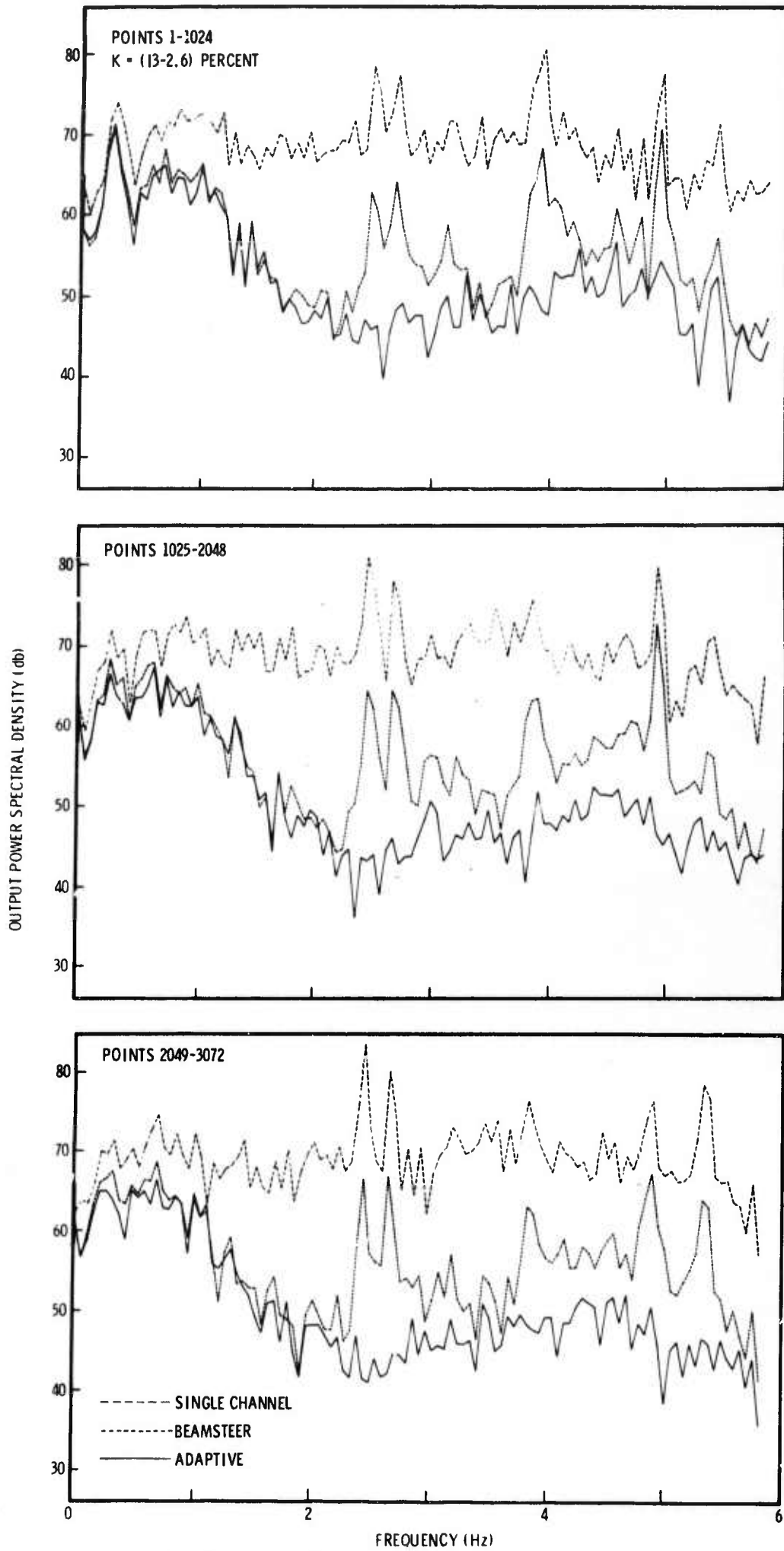Figure III-10. Impulse Responses of the 2.6 Percent Adaptive Filter

services group

Figure III-11. Improvement of the (13-2.6) Percent Adaptive Filter

services group

Figure III-12. Output Power Spectra for the (13-2.6) Percent Adaptive Filter
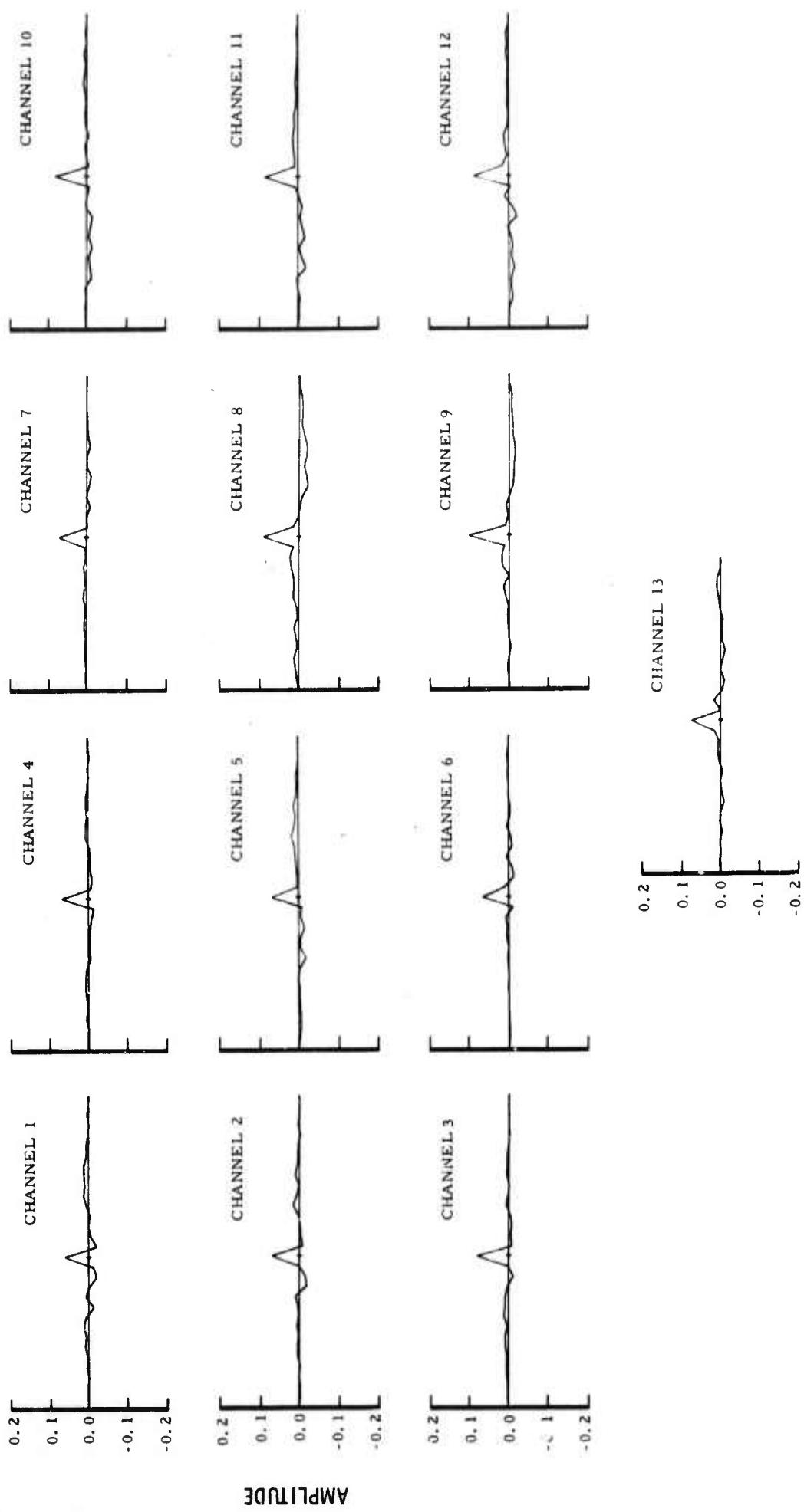
services group

Figure III-13. Impulse Responses of the (13-2.6) Percent Adaptive Filter
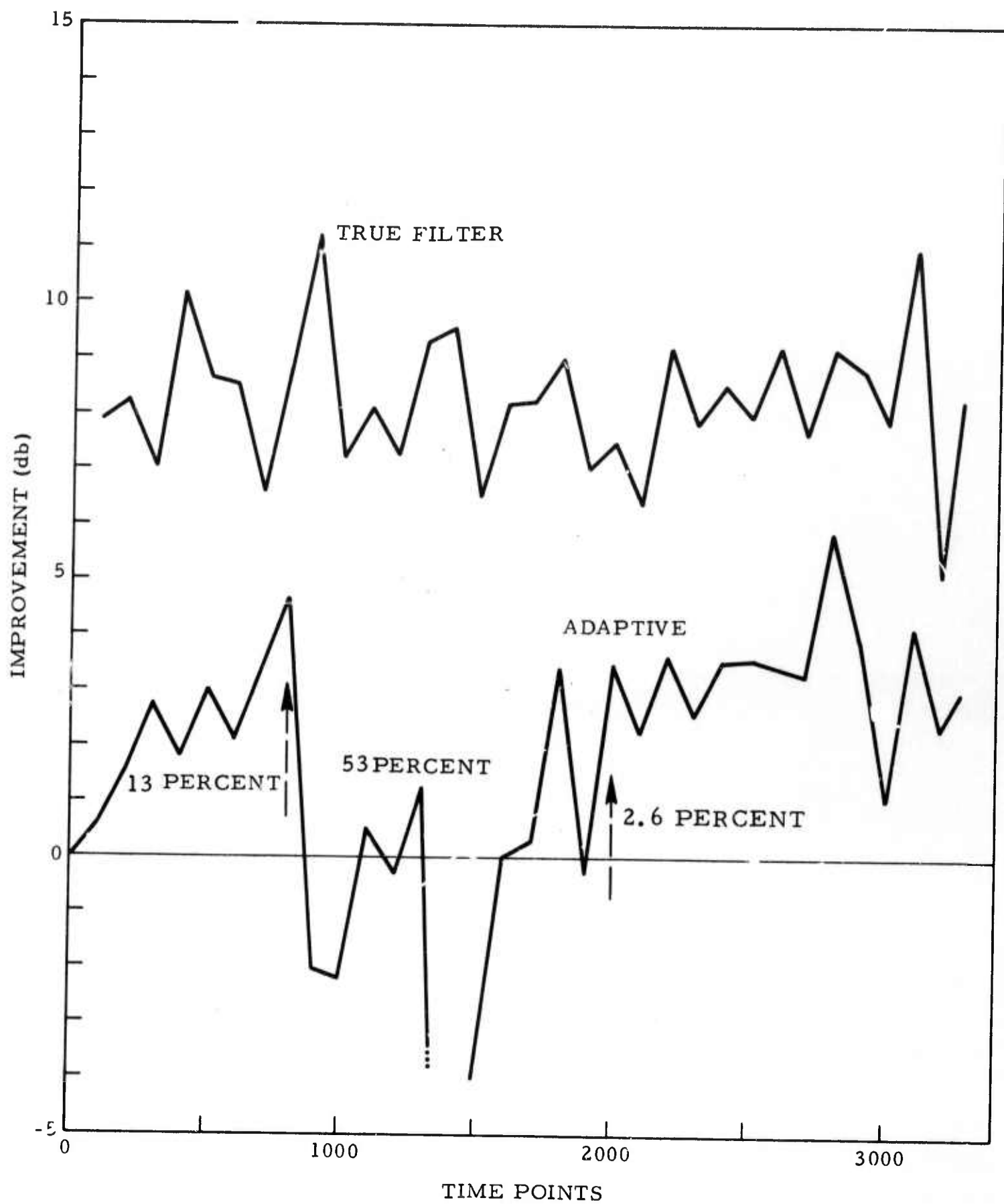
services group

Figure III-14. Improvement of the (13-53-2.6) Percent Adaptive Filter

services group

Figure III-15.   Output Power Spectra for the (13-53-2.6)
Percent Adaptive Filter

services group

Figure III-16. Impulse Responses of the (13-53.2.6) Percent Adaptive Filter.

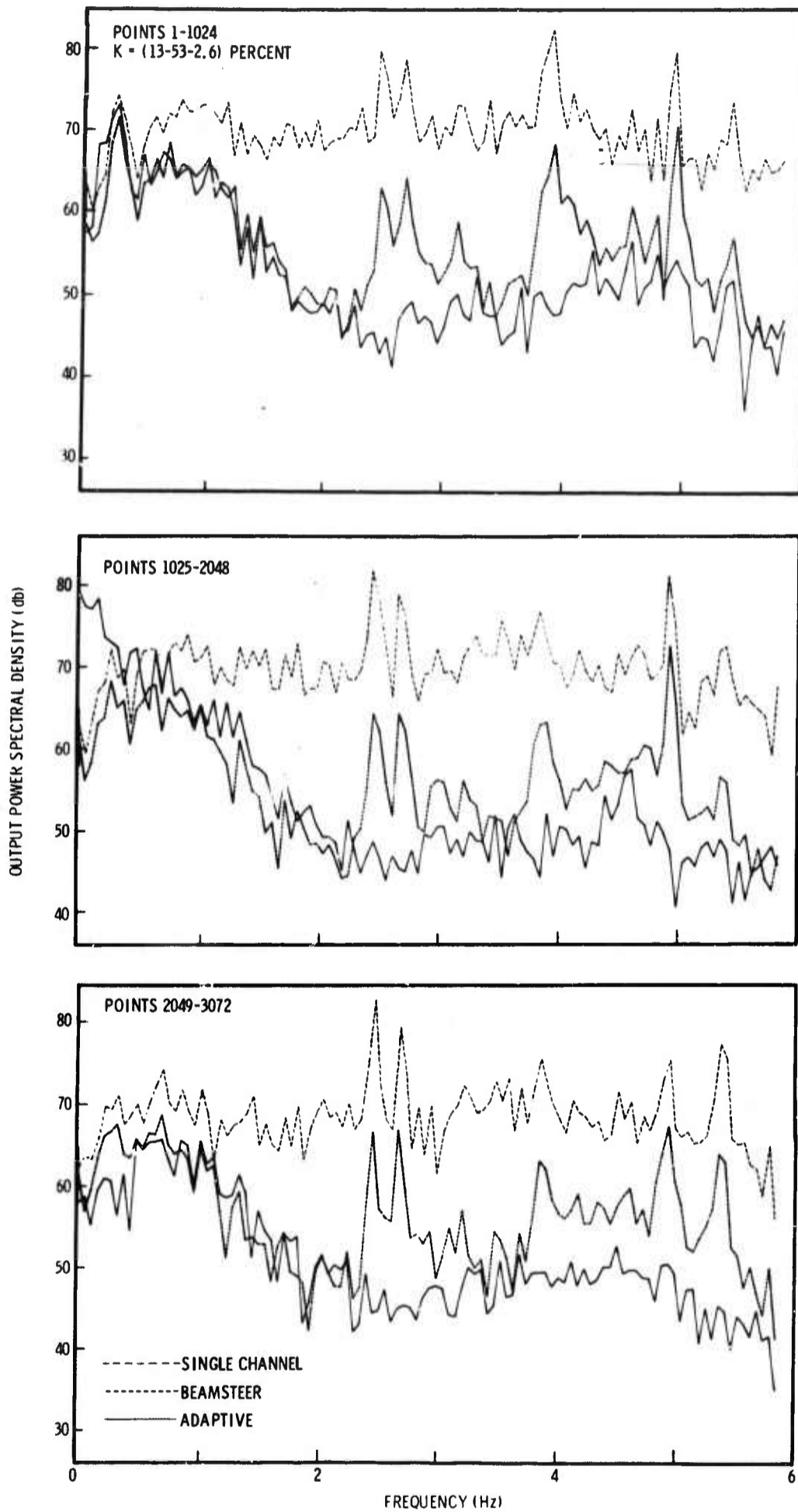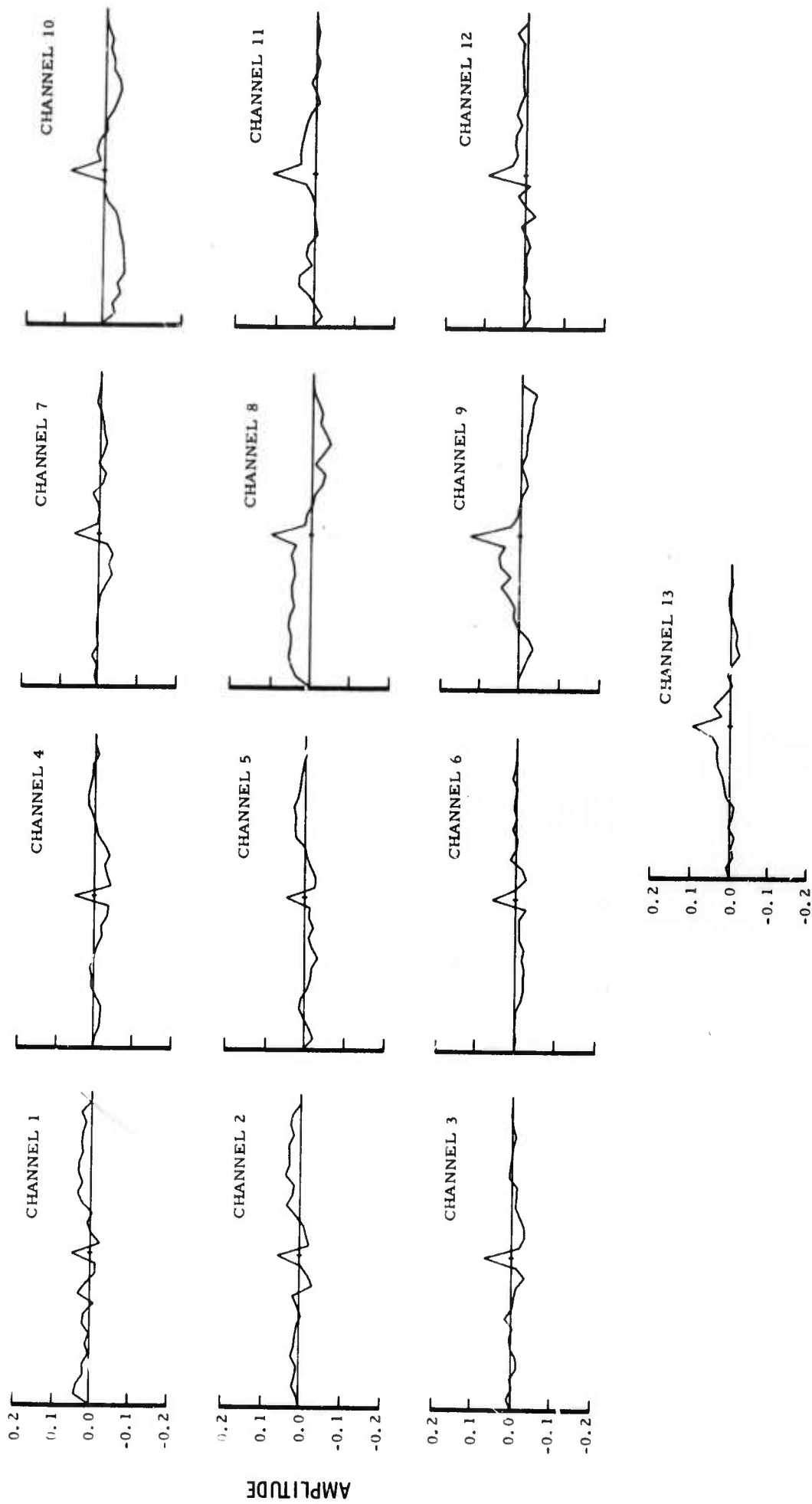Figure III-17. Improvement of the (53-13) Percent Adaptive Filter

services group
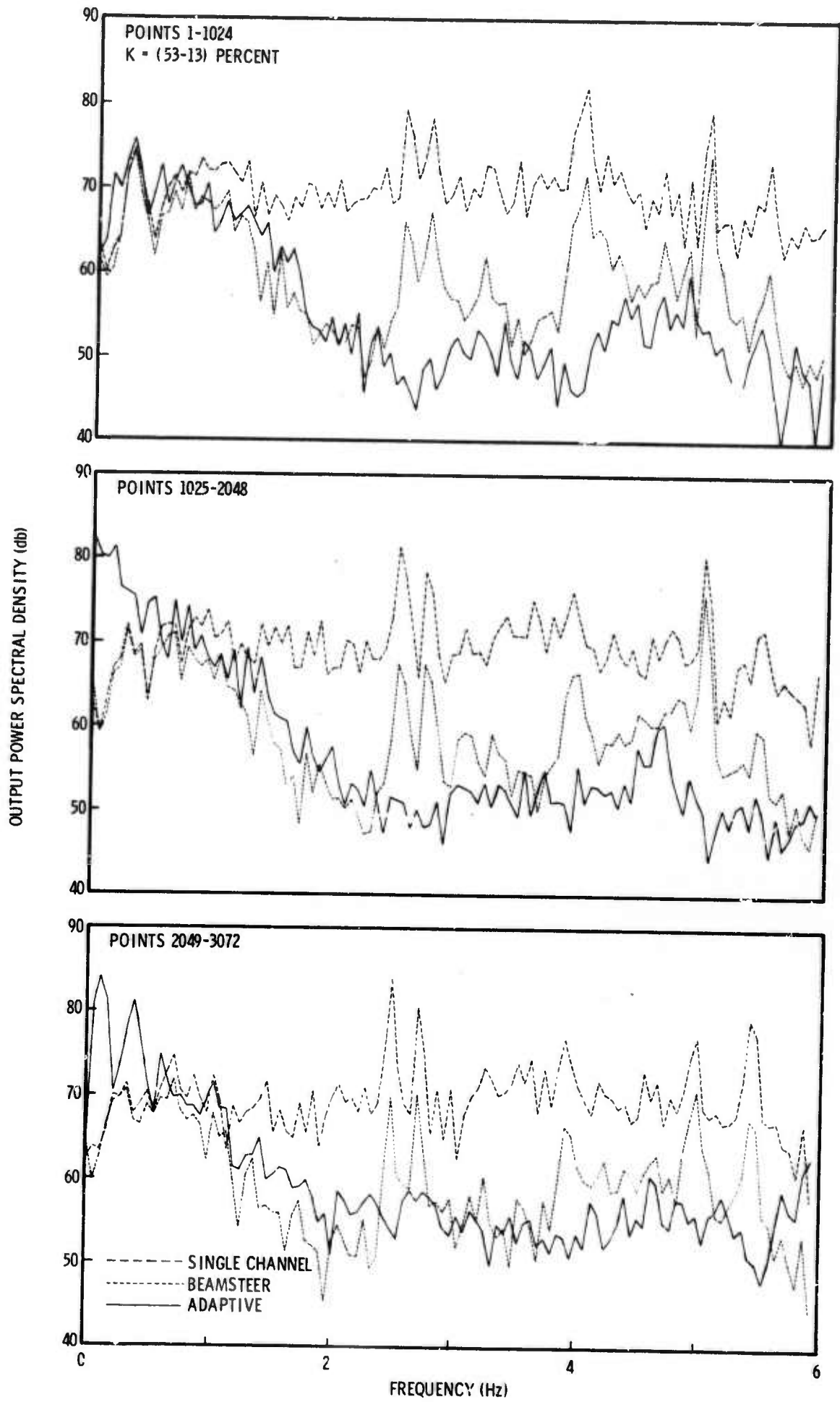
Figure III-18.   Output Power Spectra for the
(53-13) Percent Adaptive Filter

services group

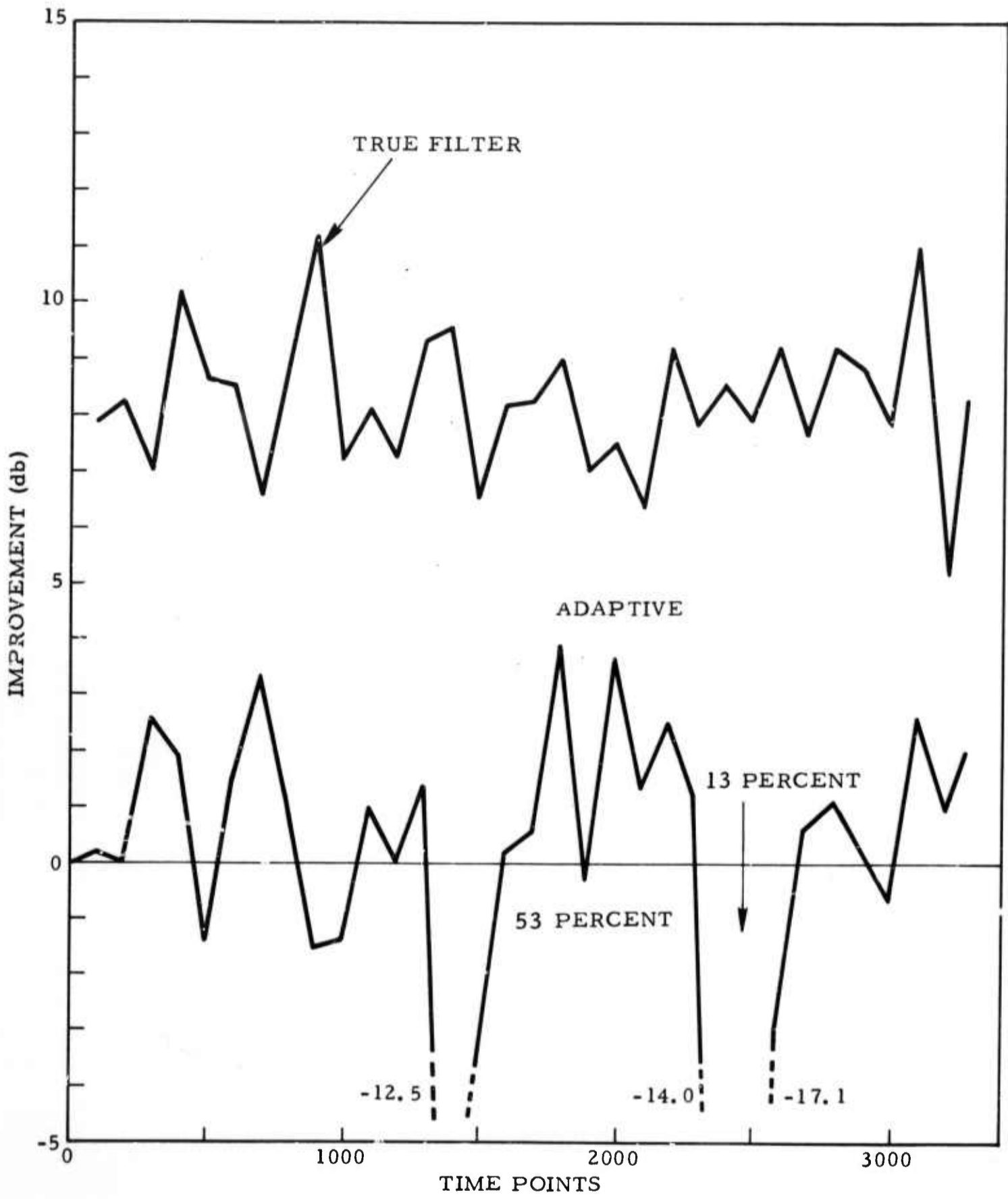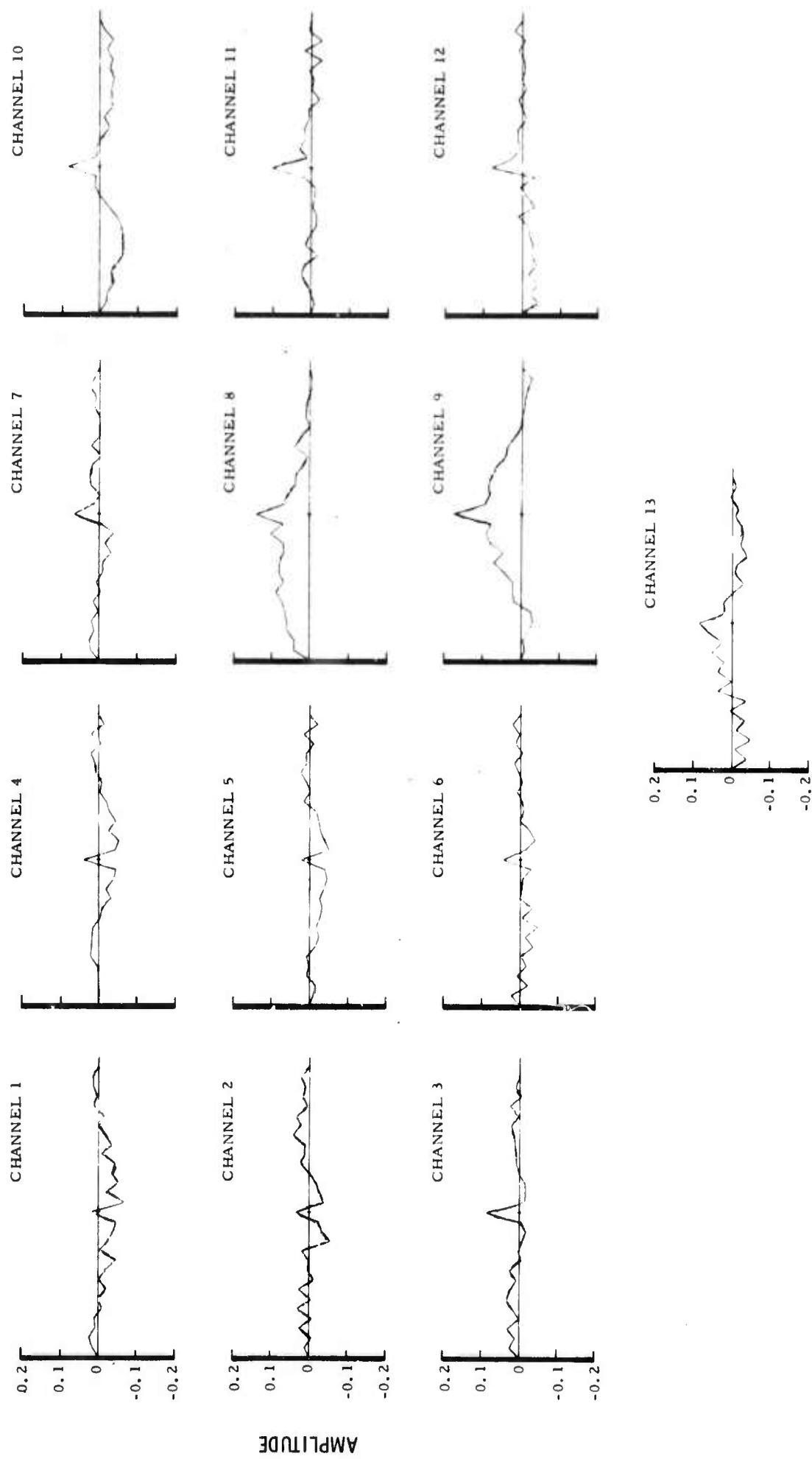Figure III-19. Impulse Responses of the (53-13) Percent Adaptive Filter

services group

Figure III-20.   Improvement of the True Filter
Adapting at a 13 Percent Rate

Figure III-21. Output Power Spectra of the True Filter
Adapting at a 13 Percent Rate

services group

Figure III-22. Improvement of the True Filter Adapting
at a 2.6 Percent Rate

Figure III-23. Output Power Spectra of the True Filter
Adapting at a 2.6 Percent Rate

services group

SECTION IV
CONCLUSIONS

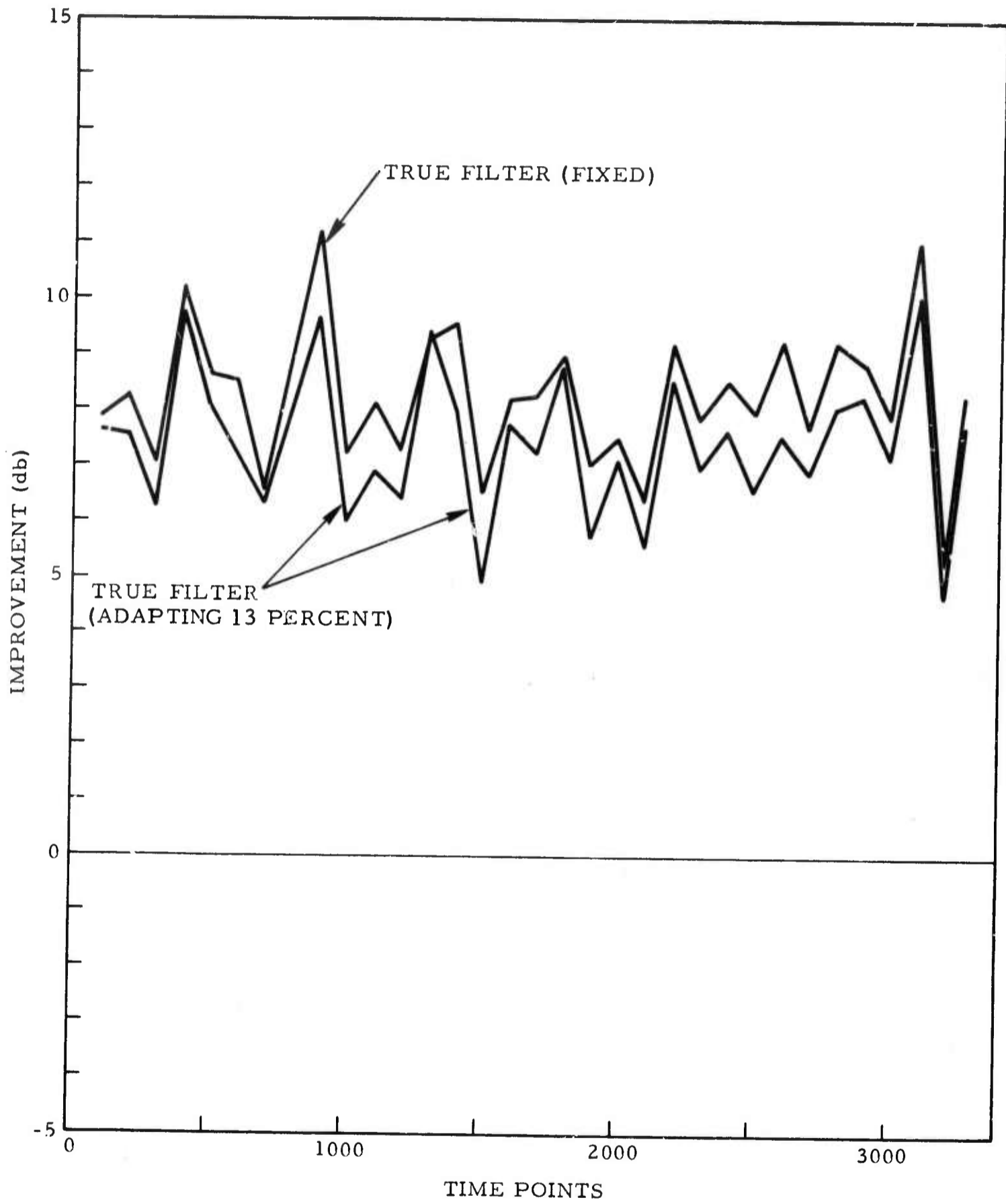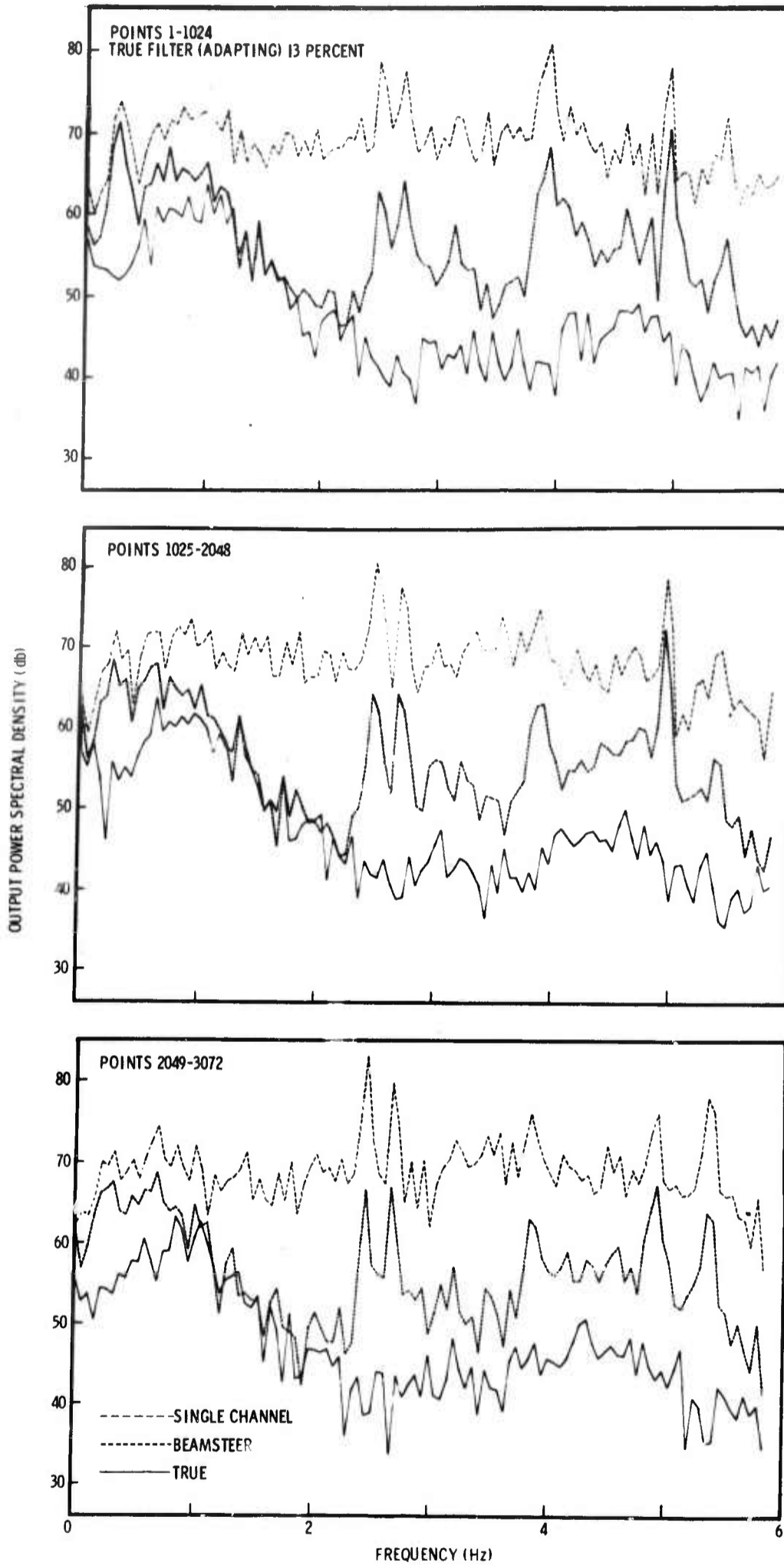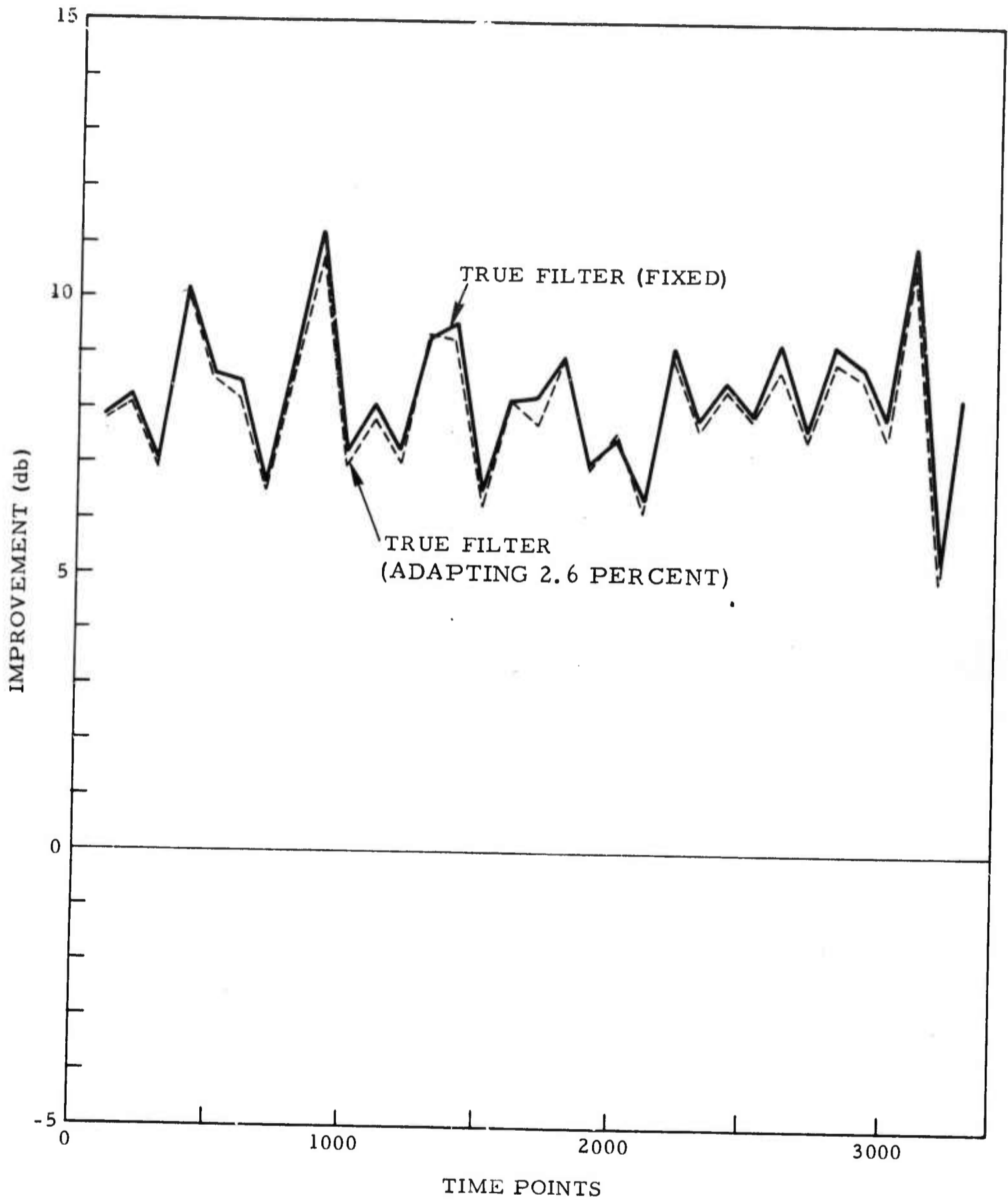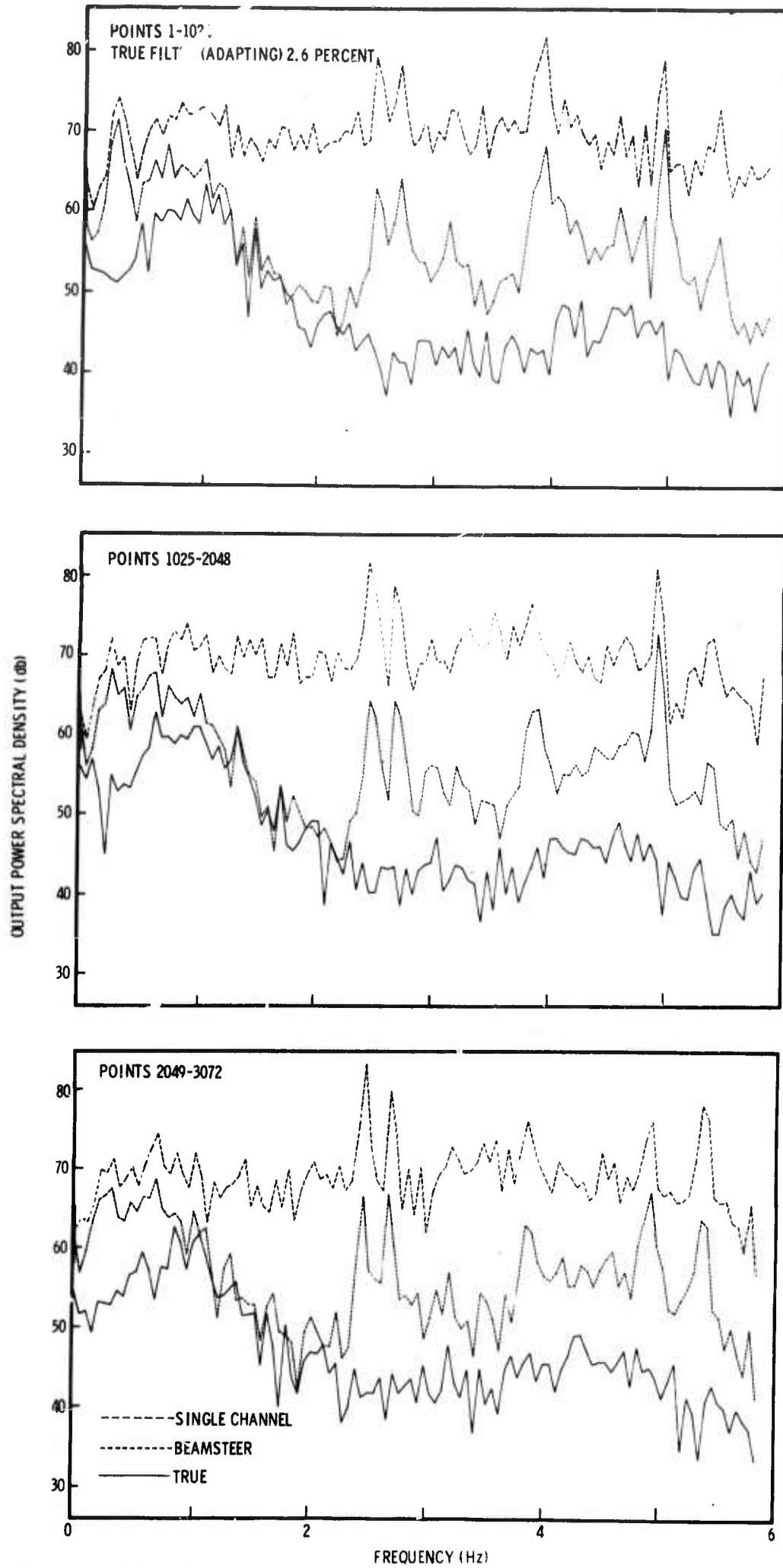The major conclusions which can be drawn from the results of this experiment are:

- The adaptive algorithm produced a quasi-equilibrium which was not but was approaching the true filter. Complete convergence was not obtained because the time constants of the process were much longer than the data available.

- The value of the convergence parameter K which produced the best adaptive filter in the least amount of time was around 13 percent of K_{max}.

- Values of K less than 3% perpeptibly slowed adaption while K greater than 50 percent distorted the spectrum of the output.

- The gradient measurement noise estimates for this data agreed well with the results predicted by existing theory.

Power from the output of the adaptive filter, for values of K less than 53 percent of K_{max}, exhibited a plateau effect where the power first decreased and then leveled off to a more-or-less constant value of 4 db less than the beamsteer power. On the same data, the true filter obtained a power 8 db smaller than the beamsteer power. The apparent failure to converge was not actually a failure but only the result of rather long time constants inherent in the data. The combination of slow adaption rates and a limited amount of data prevented attainment of time equilibrium. Attempts to speed up convergence by using larger values of the convergence parameter K, or combination of values, were not particularly successful and generally did no better than a straight value of 13 percent of K_{max}. Large values of K distorted the output spectrum while small values gave much slower convergence.

services group

The values of K which produced good results were between more than 2.6 percent and less than 53 percent of K_{max}, with the best results being obtained with a 13 percent K. The 2.6 percent value was perceptibly slower than the 13 percent rate without giving any better results. No wideband noise reduction was produced by the 53 percent rate, yet it did give good suppression of the line components and high-frequency noise. Best values of K for good noise rejection which maintain a reasonable adaption rate would be around 10 percent to 20 percent of K_{max}. After convergence, however, a smaller K would be preferable in order to reduce the amount of misadjustment. The final value of K for steady operation would require knowledge of the expected non-stationarity of the data, sampling or update rate, and tolerable misadjustment.

It must be pointed out that although the data used were synthetic, the spectrum and frequency wavenumber structure of the data can be considered typical of real (prewhitened) data. Even so, the large true filter weights indicate that the low-frequency rejection may be anomalous and attribute to false gain arising from gain inequalities. If this is true, then the optimum filters for gain-equalized data would probably have smaller filter weights, and the adaptive filters would more quickly converge. However, if the equilibrium weights are large for any reason, the adaptive algorithm may take an extremely long time to converge. This may be true, for example, at frequencies where the array aperture is equal to or smaller than the wavelength of propagating energy.

The convergence characteristics of the adaptive filter operating on-line or seismic data are still unknown. This answer awaits some actual on-line experience. This data does point out a danger in trying to design fixed filters using the adaptive algorithm and a relative short (about 4 minutes) noise sample. For such applications the method of estimating the noise correlation structure and designing filters from the correlations could give (depending on noise correlation structure) significantly better performance.

SECTION V

REFERENCES

1. Texas Instruments Incorporated, 1966: Matrix-Multiply Detection Processing of Array Data, Array Research Spec. Rpt. 15, AFTAC Contract AF 33(657)-12747, 24 May.

2. Capon, J., R. J. Greenfield, and R. J. Kolker, 1967, Multidimensional maximum-likelihood processing of a large aperture seismic array: Proceedings of the IEEE, v. 55, n. 2, p. 192, Feb.

3. Texas Instruments Incorporated, 1967: Statistics Governing the Design and Performance of Noise-Prediction Filters, Advanced Array Research Spec. Rpt. 3, AFTAC Contract F33657-67-C-0708-P001, 8 Sept.

4. Texas Instruments Incorporated, 1968: Advanced Array Research Final Report, AFTAC Contract F33657-67-C-0708-P001, 15 Feb.

5. Texas Instruments Incorporated, 1967: Adaptive Filtering of Seismic Array Data, Advanced Array Research Spec. Rpt. 1, AFTAC Contract F33657-67-C-0708-P001, 18 Sept.

6. Widrow, Bernard, 1966: Adaptive Filters I: Fundamentals, Stanford Electronics Laboratories, Stanford University, Tech. Rpt. 6764-6, Contracts DA-01-021 AMC-90015(Y) and NOBsr-95036, Dec.

7. Texas Instruments Incorporated, 1966: A Study of MCF Systems Incorporating Statistical Gain Fluctuations Designed for Cumberland Plateau Observatory, Array Research Spec. Rpt. 9, AFTAC Contract AF 33(657)-12747, 25 Mar.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Texas Instruments Incorporated Services Group P.O. Box 5621, Dallas, Texas 75222	Unclassified 2b. GROUP

3. REPORT TITLE

CONVERGENCE OF TIME-DOMAIN ADAPTIVE MAXIMUM-LIKELIHOOD FILTERS FOR STATIONARY DATA — SEISMIC ARRAY PROCESSING TECHNIQUES TECH. RPT. NO. 3

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Technical

5. AUTHOR(S) *(First name, middle initial, last name)*

William H. Swindell

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
26 February 1970	53	

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
F33657-70-C-0100	
b. PROJECT NO. VELA/T/0701/B/ASD	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. DISTRIBUTION STATEMENT

This document is subject to special export controls and each transmittal to foreign governments or foreign nationals may be made only with prior approval of Chief, AFTAC.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
ARPA Order No. 624	Advanced Research Projects Agency Department of Defense The Pentagon, Washington, D.C. 20301

13. ABSTRACT

The behavior of an adaptively designed time-domain maximum-likelihood multichannel filter during convergence on stationary data was examined. The covariance matrices of a measured seismic short-period prewhitened noise sample were used to generate 3300 time points of 13 channel stationary Gaussian data having the measured correlation structure. Using these data, 29-point adaptive filters were computed and applied. Their performance was evaluated as a function of time and compared with the performances of the beamsteer filter and the maximum-likelihood filter generated from the measured matrices.

Beginning with a beamsteer weighted initial filter, the filter was adapted for 3272 points. After 1000 adaptions, the adaptive filter was equally effective as the optimum filter in rejecting high-frequency noise. After 3272 adaptions, low-frequency noise rejection by the adaptive filter was much poorer than that of the optimum filter and not appreciably better than that of the beamsteer filter. Wideband noise reduction obtained by the best adaptive filter was about 3.5 db worse than optimum. The loss in performance was probably caused by incomplete convergence. Estimates of the gradient measurement noise are in good agreement with those predicted by theory.

DD FORM **1473**
1 NOV 65

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Seismic array processing techniques						
Time-domain adaptive maximum-likelihood filters						