XII. DIGITAL SIGNAL PROCESSING

Academic and Research Staff

Prof. Alan V. Oppenheim	Dr. Wolfgang Mecklenbraucker*
Dr. Russell M. Mersereau	Dr. Siamak Samsam†

Graduate Students

Ronald H. Frazier	Anthony P. Holt	Elliot Singer
David B. Harris	Gary E. Kopec	José M. Tribolet
	Michael R. Portnoff	

A. ENHANCEMENT OF SPEECH BY ADAPTIVE FILTERING

U. S. Navy Office of Naval Research (Contract N00014-75-C-0951)

Ronald H. Frazier, Siamak Samsam,
Louis D. Braida, Alan V. Oppenheim

1. Introduction

In a variety of situations the problem of enhancing speech degraded by the presence of a competing speaker or background noise arises. A possible key to such enhancement lies in the quasi-periodic nature of the speech waveform which corresponds to narrow harmonically spaced bands of energy in the frequency domain. One approach to speech enhancement has been to utilize a time-variant digital comb filter for which the frequency spacing of the filter passbands varies with the fundamental frequency of the speech signal that is to be enhanced.[1] When the fundamental frequency varies sufficiently slowly, the use of a comb filter leads to significant enhancement of the desired speaker, but it degrades when the fundamental frequency varies rapidly. The procedure discussed here involves the use of an adaptive filter. When the fundamental frequency is constant, this adaptive filter reduces to a comb filter but more generally takes into account the variation of fundamental frequency.

2. Adaptive Filter

To introduce the principle of the adaptive filter,[2] we consider first the implementation of a time-variant comb filter. Figure XII-1 shows a portion of a speech waveform with constant period T, on which the impulse response of a finite impulse response comb filter is superimposed. With the pitch period constant as indicated, we see that the output of the filter resulting from the desired speech signal will be the weighted sum of

*Visiting Scientist, Philips Research Laboratories, Eindhoven, The Netherlands.
†Visiting Scientist, Arya Mehr University, Iran.

corresponding points on successive pitch periods. Hence, in forming the filter output, successive periods from the desired speech will add constructively, whereas the output caused by background noise or a competing speaker with a different fundamental frequency will not. The limitation of the use of a comb filter is indicated in Fig. XII-2 where the same comb filter impulse response is applied over an interval with changing fundamental frequency. It is clear that in this case the more variation in the fundamental frequency, the less the individual pitch periods will add constructively. An alternative method is to adjust the spacing of the filter weights to coincide with the spacing of the individual pitch periods as indicated in Fig. XII-3. Such a filter no longer corresponds to a comb filter but reduces to a comb filter when the fundamental frequency is constant. It is clear, however, that when the fundamental frequency varies a comb filter is less desirable than the adaptive filter of Fig. XII-3.

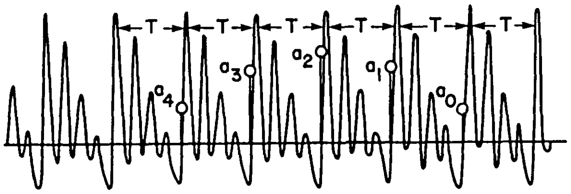

Fig. XII-1.
Comb filter and speech waveform with constant pitch period.

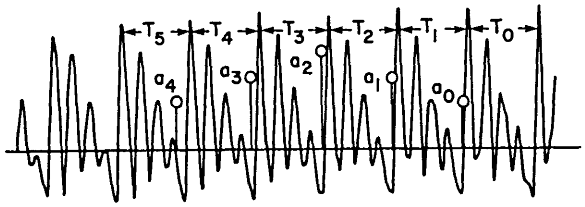

Fig. XII-2.
Comb filter and speech waveform with nonconstant pitch period.

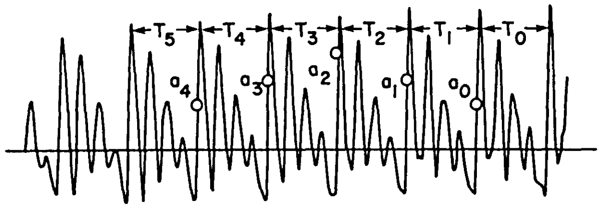

Fig. XII-3.
Adaptive filter and speech waveform with nonconstant pitch period.

Another manner of viewing the adaptive method is indicated in Fig. XII-4. Consider the waveform broken into segments according to the pitch epochs and then aligned as shown. A weighted average is computed point by point as the filter moves in the indicated direction. From an intuitive viewpoint this operation computes an average unit

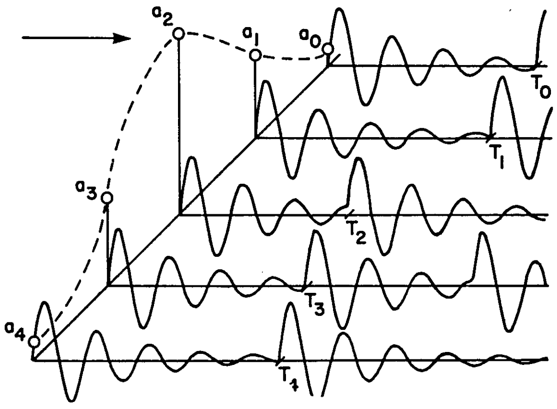

Fig. XII-4. Adaptive filter, segmented view.

sample response based on the several previous periods. This technique works well in conjunction with the assumption that the impulse response of the vocal tract is slowly varying. From the separational aspects this procedure allows the components from the desired speaker to be added coherently while the contributions from the undesired speaker are added incoherently.

3. Overload Problem with Correction

One difficulty with the adaptive procedure is illustrated in Fig. XII-5. In normal voiced speech there are some areas where the pitch period changes very rapidly in a short time interval. This phenomenon creates segments on the speech waveform that are much shorter than neighboring segments. This is illustrated in Fig. XII-5 by the short segment that is terminated at T_2. As the filtering moves to the right, computing

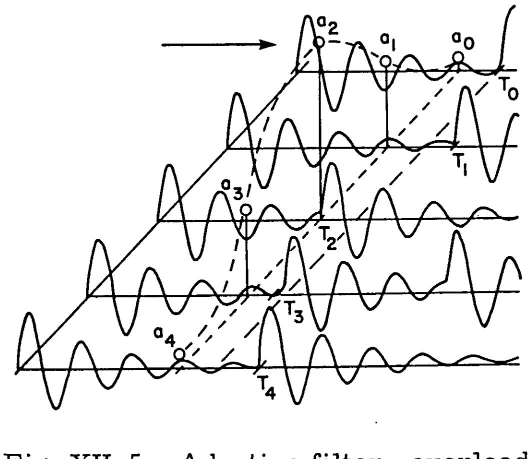

Fig. XII-5. Adaptive filter, overload problem.

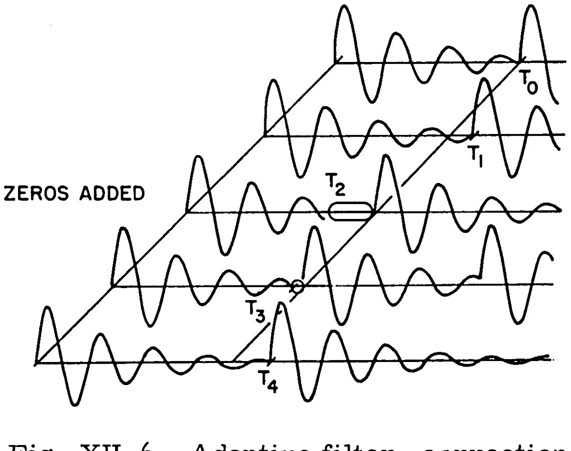

Fig. XII-6. Adaptive filter, correction of overload problem.

a point-by-point average, no problems arise before point T_2. As the coefficient a_2 moves past point T_2, the procedure disagrees with the adaptive filtering concepts. A possible solution to this problem is displayed in Fig. XII-6. If the filter is being controlled by the segment terminated at T_0, then the proposed method is to lengthen the short segments to correspond to the control segment. This is done by padding the short segments with zeros as shown. This procedure may be thought of as a "turning off" of the coefficients that are involved in short segments. The actual computer implementation performed additional operations in order to prevent fluctuations in gain of the output waveform.

4. Results

In order to analyze the adaptive filter performance before an actual speech waveform was processed, a test signal which was "speechlike" in form was processed. This test signal was generated from a damped sine wave that was convolved with a nonuniformly spaced train of unit samples. The spacing between the samples was prepared to vary about a mean spacing. This test waveform served as a good model for the speech waveform, and the impulse response and pitch period were known exactly. Figure XII-7 demonstrates the capability of the adaptive system on the test input signal. For this case the filtering system is prepared to act as an identity system in order to illustrate the amount of desired speaker distortion induced by the systems. The input signal is shown in Fig. XII-7a, while the outputs from the comb and adaptive filters are displayed in Fig. XII-7b and 7c, respectively.

With an actual speech waveform, the comparisons between systems cannot as easily

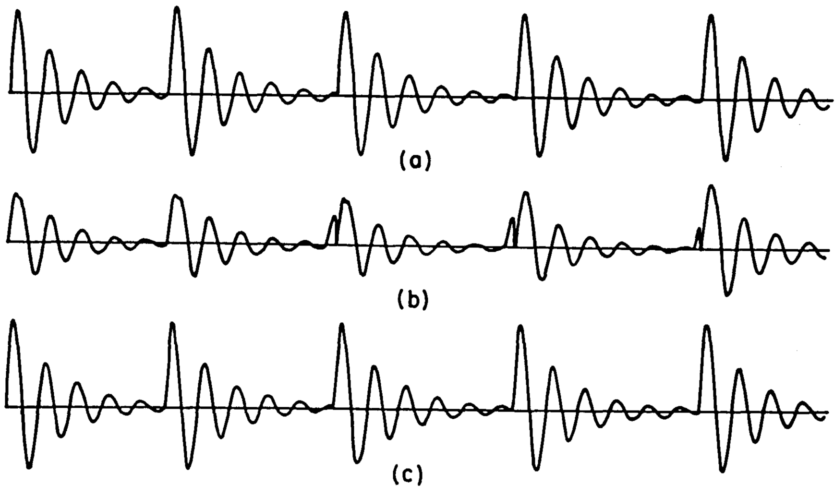

Fig. XII-7. (a) Test signal input waveform.
(b) Comb filter output waveform, identity system.
(c) Adaptive filter output waveform, identity system.

be made. From informal listening and spectrographic analysis, a limited amount of evaluation was conducted. The adaptive filter definitely provides enhancement of speech degraded by the presence of competing speakers and background noise. Some potential improvements are yet to be investigated with regard to handling voiced-unvoiced transitions, optimum impulse response length, etc. Furthermore, we did not investigate the complex question of pitch detection. A simple procedure of measuring the glottal pressure waveform with an accelerometer[3] to obtain pitch period information before the two speakers were added was employed, but any error in the pitch epoch marking would also introduce some distortion in the output waveform.

These issues and the entire question of system performance will be examined in future work with extensive listening tests.

References

1. V. C. Shields, Jr., "Separation of Added Speech Signals by Digital Comb Filtering," S. M. Thesis, Department of Electrical Engineering, M. I. T., September 1970.
2. R. H. Frazier, "An Adaptive Filtering Approach toward Speech Enhancement," E.E. Thesis, Department of Electrical Engineering, M. I. T., June 1975.
3. W. L. Henke, "Signals from External Accelerometers during Phonation: Attributes and Their Internal Correlates," Quarterly Progress Report No. 114, Research Laboratory of Electronics, M. I. T., July 15, 1974, pp. 224-231.

B. AMPLITUDE FILTERING USING A TIME-VARIABLE TRANSFORMATION

U. S. Navy Office of Naval Research (Contract N00014-75-C-0951)
National Science Foundation (Grant ENG71-02319-A02)

Elliot Singer, Alan V. Oppenheim

1. Introduction

In a recent paper[1] Moore and Parker described a filtering scheme for separating signals by their amplitude characteristics rather than by their frequency spectra. The filter, called an E-filter, distorts the time variable of the signal so as to map segments with different amplitude characteristics into different frequency bands even if their original spectra overlap. One application for such a filter is the removal of additive noise from a broadband signal without smoothing sharp edges in the signal. Alternatively, the system could be used as a selective peak filter, passing the high-amplitude signal peaks while attenuating the low-amplitude peaks. We report here on a project that analyzed and evaluated the E-filter. One aspect has been the determination of an alternative implementation of the E-filter which lends itself more easily to available analytic techniques. Although the alternative scheme is shown to be less practical than the original, its derivation clarifies some of the issues surrounding E-filters.

2. E-Filters

A representation of the E-filter proposed by Moore and Parker is shown in Fig.XII-8.

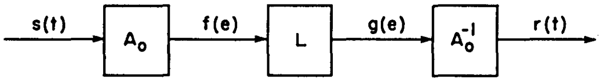

Fig. XII-8. E-filter proposed by Moore and Parker.[1] Note that g(e) is the equivalent of Moore and Parker's label $f^*(e)$.

The system A_o performs a transformation that "replots" the input signal s(t) against a new "time" variable e, where e is the arc length accumulated by s(t) up to time t. Thus $A_o\{s(t)\}$ is defined by the relation

$$f(e)\big|_{e=\theta(t)} = s(t), \tag{1a}$$

where

$$\theta(t) = \int_0^t \sqrt{1 + \dot{s}^2(\tau)}\, d\tau \tag{1b}$$

and θ(0) = 0. The transformation performed by A_o is illustrated in Fig. XII-9. The system L with input f(e) and output g(e) is modeled as an ideal lowpass filter. The signal g(e) is the input to the system A_o^{-1} which produces the final output r(t).

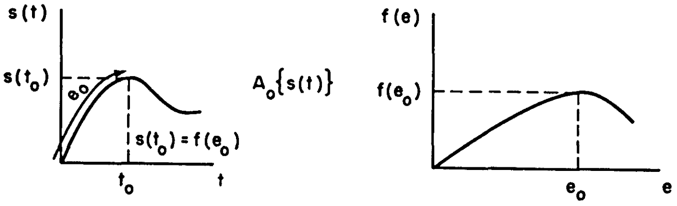

Fig. XII-9. Transformation performed by A_o.

A_o^{-1} corresponds to the implementation used by Moore and Parker to invert the input transformation A_o; that is, $A_o^{-1}\{g(e)\}$ is defined by the relation

$$r[\theta^{-1}(e)] = g(e) \tag{2a}$$

or

$$r(t) = g(e)\big|_{e=\theta(t)}. \tag{2b}$$

As shown in Fig. XII-10 the mapping used by A_o to transform from t to e is now reversed so as to produce a value t for every corresponding value of e. Thus the output r(t) of the E-filter is the function that when plotted against the arc length of s(t) produces g(e). It is useful to interpret the transformations performed by the systems A_o and A_o^{-1} as instantaneous time warpings of their input signals. Thus $f(e) = f[\theta(t)]$ corresponds to a time-expanded version of s(t) with an instantaneous expansion given by

$$\dot{\theta}(t) = \sqrt{1 + \dot{s}^2(t)}. \tag{3}$$

After being lowpass filtered, the signal is compressed by the system A_o^{-1} in a manner that corresponds exactly in time to the instantaneous expansion performed by the system A_o. This is easily shown by recognizing that the Moore and Parker implementation of the inverse transformation A_o^{-1} requires that

$$t = \theta^{-1}[\theta(t)] \tag{4}$$

and thus

$$\dot{\theta}^{-1}(e) = \frac{1}{\theta(t)}, \tag{5}$$

where the reciprocal of $\dot{\theta}^{-1}(e)$ corresponds to the instantaneous compression factor. Consequently, if the impulse response of the lowpass filter has a short duration relative to that of f(e), then the duration of r(t) will be virtually identical to that of s(t).

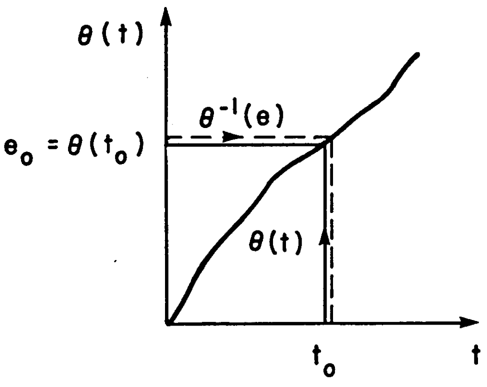

Fig. XII-10.
Mapping reversal used by Moore and Parker in the inverse system A_o^{-1}.

Maintaining an exact temporal correspondence between the signal durations of the input and the output of the E-filter is a fundamental objective of the Moore and Parker system. Since a linear time-invariant filter alone cannot separate signal components whose spectra overlap, a nonlinear time warping is performed to divide the signal spectrally according to its high and low amplitude components. Linear filtering can then be employed to extract the desired component. In returning the resulting signal to the t-domain, processing is performed by A_o^{-1} to invert the time warping applied earlier. The disadvantage of the Moore and Parker approach to the inversion problem is that it requires that the transformation performed by A_o be saved and made available to A_o^{-1}. Hence the system representation of Fig. XII-8 is misleading in that it implies that successive stages of the system are independent of the operations performed previously. A more explicit representation of their system is shown in Fig. XII-11 which clearly indicates a channel between A_o and A_o^{-1} containing the time warping transformation.

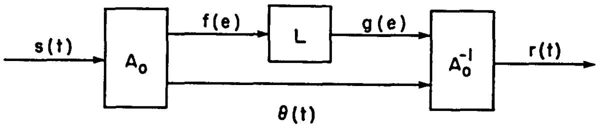

Fig. XII-11. Explicit representation of the system of Moore and Parker.

3. Homomorphism and an Alternative System

An unresolved issue discussed by Moore and Parker is the question of whether or not the E-filter is homomorphic.[2] The difficulty arises essentially because their implementation of the E-filter is a connection of interdependent subsystems and thus the ability to determine the invertibility of the subsystems is lost. Consequently, the problem of finding a homomorphic implementation of the E-filter reduces to the problem of determining a representation composed of independent and invertible subsystems, which in turn implies finding a single-input single-output inverse system. A schematic representation of this system would have the form shown in Fig. XII-12, where T is identical to the system A_o but has only a single output channel and T^{-1} is the single-channel inverse of T. The single-channel inverse transformation T^{-1} can be derived by considering carefully the nature of the forward transformation. The system T (or A_o) produces f(e) by replotting its input s(t) against its accumulated arc length. The inverting system T^{-1} with input g(e) must therefore have an output r(t) which when plotted against its arc length gives g(e), as shown in Fig. XII-13. Then

$$(de)^2 = (dg)^2 + (dt)^2. \tag{6}$$

Therefore

$$t = \theta_g^{-1}(e) = \int_0^e \sqrt{1 - \dot{g}^2(\alpha)}\, d\alpha. \tag{7}$$

The subscript in $\theta_g^{-1}(e)$ indicates explicitly the dependence on the function g(e).

Thus the single-channel system can be described by the relations

$$f(e)\big|_{e=\theta_s(t)} = s(t), \tag{8a}$$

where

$$\theta_s(t) = \int_0^t \sqrt{1+\dot{s}^2(\tau)}\ d\tau \tag{8b}$$

and

$$r(t) = g(e)\big|_{e=\theta_g(t)}, \tag{9a}$$

where

$$\theta_g^{-1}(e) = \int_0^e \sqrt{1-\dot{g}^2(\alpha)}\ d\alpha. \tag{9b}$$

It is important to note that while the single-channel system is guaranteed to be homomorphic, it does not have the property of preserving the correspondence between its input and output signal durations. This is due to the fact that the instantaneous compression factor associated with T^{-1},

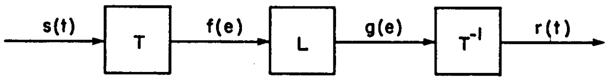

Fig. XII-12. Single-channel representation of an E-filter.

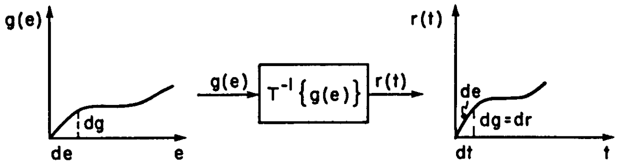

Fig. XII-13. Single-channel inverse T^{-1}.

$$\frac{1}{\dot{\theta}_g^{-1}(e)} = \frac{1}{\sqrt{1-\dot{g}^2(e)}}, \tag{10}$$

is no longer dependent strictly on f(e). This compression factor is plotted as a function of $|\dot{g}(e)|$ in Fig. XII-14. Since g(e) is obtained from f(e) by lowpass filtering, g(e) will have a slope whose rate of change is slower relative to that of f(e). It is evident that

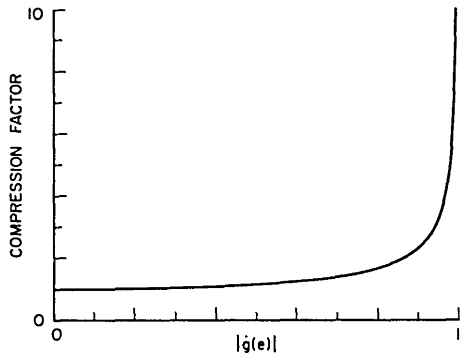

Fig. XII-14.
Compression factor vs $|\dot{g}(e)|$.

smoothing affecting values of $|\dot{g}(e)|$ near unity could easily have a large effect on the compression factor, with a resulting increase in the duration of r(t) relative to that of s(t).

4. Evaluation

In order to evaluate the performance of the E-filter, a series of simulations of both the Moore and Parker and the single-channel systems was carried out. Whenever possible, the discrete-time algorithms used by Moore and Parker to represent continuous-time operations were implemented. Thus integrals were replaced by summations and derivatives by backward first differences.

Inputs to the systems were samples of various combinations of sinusoids, square waves, and triangular waves. The sampling rate was between 40 and 80 samples per period of the high-amplitude signal. The Moore and Parker condition that the maximum amplitude of the signal be greater than one quarter of the period was also met. Since arc length was approximated by the cumulative distance between sample points, and since these values were generally nonintegers, the function $f(e_n)$ at the output of A_o and T required some form of interpolation before it could be filtered. The method chosen was simple linear interpolation, which Moore and Parker also used. The linear filter was causal with a rectangular impulse response whose length was specified by the amplitude of the high-amplitude signal. Based on the analysis of Moore and Parker this length was chosen to be one-half or less of the maximum amplitude of the input signal. The sequence of operations for both systems is illustrated in Fig. XII-15.

The performances of the Moore and Parker and the single-channel systems were compared for a series of high-amplitude signals with superimposed low-amplitude, high-frequency signals. The results of the computations are shown in Fig. XII-16, with the time distortion effect of the single-channel system clearly in evidence. It should be pointed out that in the digital implementation of the system signal smoothing is the result of interpolation, as well as of lowpass filtering, and therefore the interpolation performed in the single-channel system must be accomplished at a very high rate to minimize time distortion.

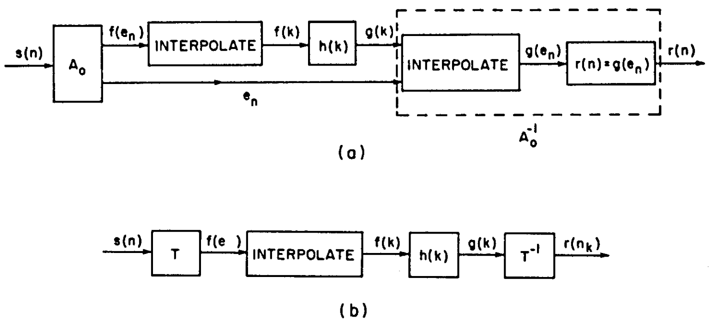

Fig. XII-15. Discrete-time implementation of E-filters.
(a) System of Moore and Parker.
(b) Single-channel system.

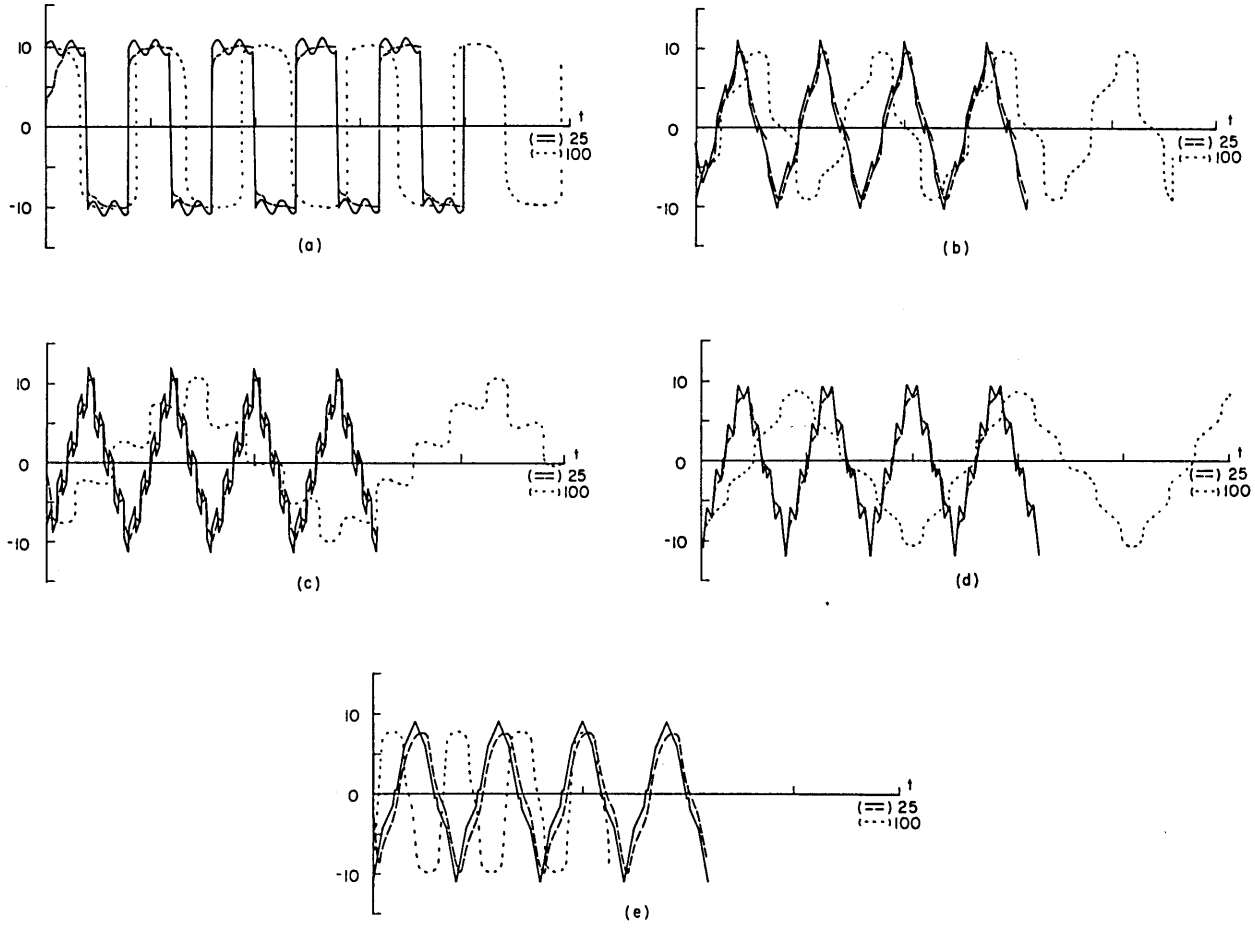

Fig. XII-16. Comparison of output performance of Moore and Parker system (---) and single-channel system (...) for a sum of signals. Note the change in scale for the single-channel system. Input: (a) square wave + sinusoid, (b) and (c) triangular wave + square wave, (d) and (e) triangular wave + triangular wave.

One of the conclusions that may be drawn from the results of the simulations is that the single-channel system has a severe disadvantage when compared with the Moore and Parker system. Any smoothing that takes place within their system will be reflected in smoothing at the output, whereas smoothing within the single-channel system is reflected in smoothing and severe time distortion at its final output. This seems to limit the usefulness of the single-channel system. As far as the Moore and Parker system is concerned, however, the results of this project show that it appears to have some value as an amplitude filter. Moore and Parker indicate that the filter has been very useful when applied to their work on pattern recognition which requires peak extraction.

References

1. D. J. H. Moore and D. J. Parker, "On Nonlinear Filters Involving Transformation of the Time Variable," IEEE Trans., Vol. IT-19, No. 4, pp. 415-422, July 1973.
2. A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, Jr., "Nonlinear Filtering of Multiplied and Convolved Signals," Proc. IEEE 56, 1264-1291 (1968).