ENHANCEMENT OF SPEECH BY ADAPTIVE FILTERING*

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

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts

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.

Adaptive Filter

To introduce the principle of the adaptive filter,[2] we consider first the implementation of a time-variant comb filter. Figure 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 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.

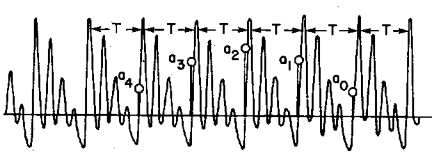

Figure 1.
Comb filter and speech waveform with constant pitch period.

The limitation of the use of a comb filter is indicated in Figure 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 Figure 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 Figure 3.

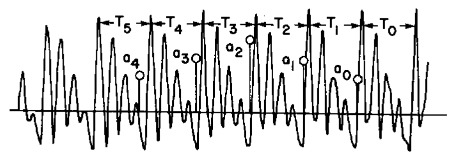

Figure 2.
Comb filter and speech waveform with nonconstant pitch period.

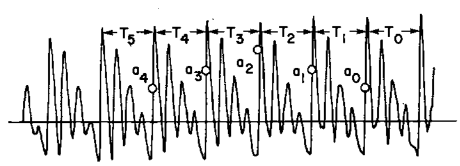

Figure 3.
Adaptive filter and speech waveform with nonconstant pitch period.

Another manner of viewing the adaptive method is indicated in Figure 4. Consider

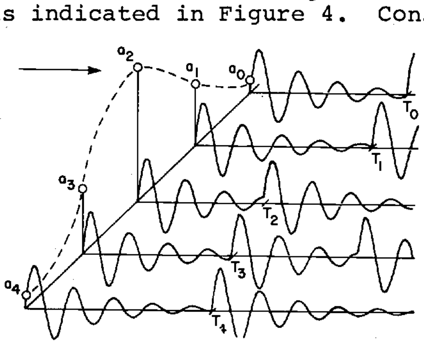

Figure 4.
Adaptive filter, segmented view.

the waveform broken into segments according to the pitch epochs and then aligned as shown. A weighted avarage is computed point by point as the filter moves in the indicated direction. From an intuitive viewpoint this

*This work was supported by the Advanced Research Projects Agency monitored by ONR under Contract N00014-75-C-0951.

†U.S. Coast Guard, Wildwood, New Jersey.

††Arya Mehr University, Teheran, Iran.

operation computes an average unit 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.

Overload Problem with Correction

One difficulty with the adaptive procedure is illustrated in Figure 5. In normal

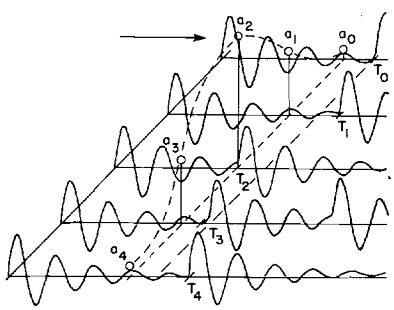

Figure 5.
Adaptive filter, overload problem.

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 Figure 5 by the short segment that is terminated at T_2. As the filtering moves to the right, computing 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 Figure 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.

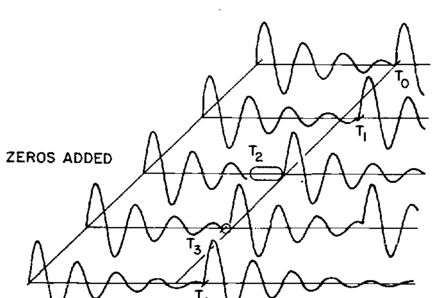

Figure 6. Adaptive filter, correction of overload problem.

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 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 Figure 7a, while the outputs from the comb and adaptive filters are displayed in Figure 7b and 7c, respectively.

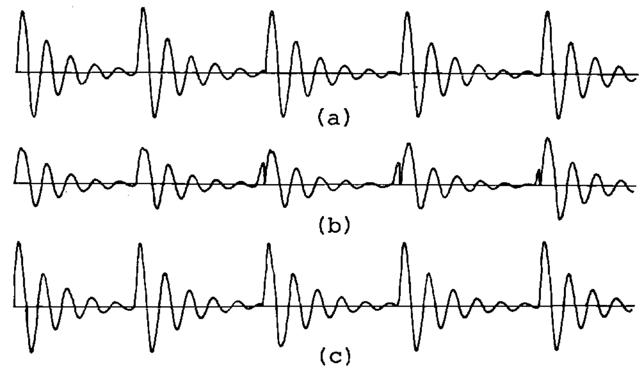

Figure 7.
(a) Test signal input waveform.
(b) Comb filter output waveform, identity system.
(c) Adaptive filter output waveform, identity system.

With an actual speech waveform, the comparisons between systems cannot as easily 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, Massachusetts Institute of Technology, September 1970.

2. R. H. Frazier, "An Adaptive Filtering Approach toward Speech Enhancement," E.E. Thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, 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, Massachusetts Institute of Technology, July 15, 1974, pp. 224-231.