"Speech Processing Applications of Switched-Capacitor Filters", Case Studies in Signal Processing, pp.14-18, IEEE, Scotland, 1979

SWITCHED-CAPACITOR APPLICATIONS IN SPEECH PROCESSING

R. W. Brodersen, P. J. Hurst and D. J. Allstot

Department of Electrical Engineering and Computer Sciences

and the Electronics Research Laboratory

University of California, Berkeley, California 94720

INTRODUCTION

Recent advances in MOS-LSI technology make it possible to perform complex signal processing functions on a single integrated circuit (IC). This increase in complexity has been made possible by not only the well known advances in MOS digital technology, but also by recent developments in MOS analog circuits . In particular, switched capacitor techniques make it possible to design stable accurate recursive filters. These new circuit techniques exploit unique characteristics of MOS integrated circuits to perform analog discrete time signal processing. In particular, the ability to store, sense and move charge (which represents signal) on capacitors is used, as well as the ability to define extremely accurate (<.1%), stable ratios of capacitor values. These ratios can be used to define such parameters as precise gains and filter frequency response characteristics. These integrated analog MOS circuits can be designed to be virtually free from many of the disadvantages of classical analog circuits, such as the temperature instability of gains and offsets as well as the need for trimming and adjustments. In addition, since the MOS analog discrete time approach uses a technology which is compatible with high density digital circuitry, a combination of analog and digital circuits can be implemented on a single IC. This gives an extra degree of freedom to the designer since he can choose the most appropriate partitioning between analog and digital techniques without the complexity and problems of interconnecting several IC's fabricated in different technologies. As a general rule, functions such as data formatting, long term data storage or simple decision making are often best performed digitally, whereas relatively complex high speed deterministic processing such as filtering and peak detection can be performed much more efficiently in the analog domain.

In order to explore this tradeoff between analog and digital processing, two specific signal processing applications will be investigated in the area of speech synthesis and analysis. Both of these applications make extensive use of switched-capacitor filters, so the basic elements of these filters will be described in the next section.

SWITCHED CAPACITOR FILTERS

A possible approach to monolithic filters would be to attempt to integrate in MOS technology conventional RC active circuits. However, this is not possible because of the necessity of accurately defining resistance capacitance products, which requires that the absolute value of the resistors and capacitors be well controlled. Typical variations in an MOS process yield variations of these parameters on the order of $\pm 20\%$. Also the area required to implement resistors and capacitors which yield RC products that are needed for audio frequencies are prohibitively large.

A circuit that performs the function of a resistor is shown in Fig. 1(a). The operation of this "resistor" is as follows: the switch is initially in the left-hand position so that the capacitor C is charged to the voltage V_1.

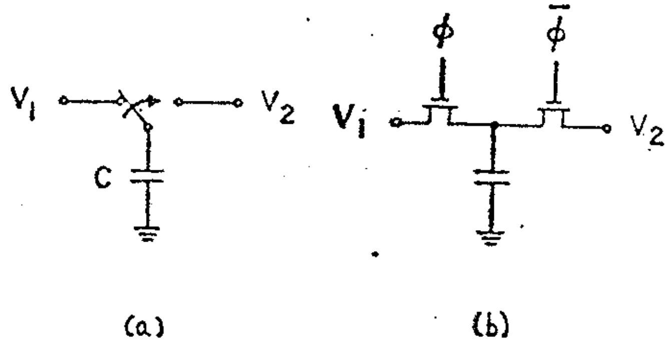

Fig. 1(a) A switched-capacitor resistor.
1(b) The MOS implementation.

The switch is then thrown to the right and the capacitor is discharged to the voltage V_2. The amount of charge which flows into is thus $Q = C(V_1 - V_2)$. If the switch is thrown back and forth every T_c seconds, then the current flow i_s into V_2 will be

$$i_s = \frac{C(V_1 - V_2)}{T_c} \qquad (1)$$

Thus the size of an equivalent resistor which would perform the same function as this circuit is $R = T_c/C$. The MOS realization of the circuit of Fig. 1(a) is shown in Fig. 1(b). The two MOSFET's are operated as switches which are pulsed with a two phase nonoverlapping clock (ϕ and $\bar{\phi}$) at a frequency f_c. The most important advantage of the switched capacitor resistors is the high accuracy of RC time constants that can be obtained with their use. If a capacitor C_1 which is switched at a clock rate of f_c is connected to a capacitor C_2, the resultant time constant of this RC network is approximately

$$\tau_{RC} \approx \frac{1}{f_c}\frac{C_2}{C_1} \qquad (2)$$

For a given clock rate the value of τ_{RC} is therefore determined by a ratio of capacitor values which makes it insensitive to most processing variations [1].

The problem of implementing active filters in MOS technology is thus reduced to the question of what kind of active filter should be used. The filter configuration chosen is the analog computer simulation of the equations which describe a passive doubly-terminated RLC ladder. These filters are called "leapfrog" or "active ladder" filters in the active filter circuit literature and are closely related to wave digital filters in the digital signal processing literature.

The basic building block of these filters is a differential integrator and summer which is shown implemented in the conventional way with resistors and capacitors in Fig. 2(a). The equations which describe this circuit are

$$V_{OUT}(\omega) = -\frac{C_S}{C_I}V_1 + \frac{1/RC_I}{j\omega}(V_3 - V_2) \qquad (3)$$

Straightforward substitution of the resistors by the circuit in Fig. 1 would yield a switched capacitor circuit that is sensitive to parasitics. So an alternate method of realizing the resistors which also makes it possible to invert the sign of the signal is used and is shown in Fig. 2(b). The circuit in this figure is described by the following z-transform equation

$$V_{OUT}(z) = -\frac{C_S}{CI}V_1 + \frac{C_u/C_I}{1-z^{-1}}(z^{-1/2}V_3 - V_2) \qquad (4)$$

Appropriate interconnection of these integrators makes it possible to realize any order filter. The leapfrog or active ladder configuration is particularly useful and design of filters using this approach has been discussed extensively [1-4].

SPEECH SYNTHESIS

There are a variety of approaches which can be taken for an integrated speech synthesizer which tradeoff the following: the complexity of the analysis to obtain the parameters to drive the synthesizer; the data rate of these parameters; and the silivon area to in-egrate the circuit.

The simplest analysis-synthesis approach is to directly encode the speech waveform without making any assumption about the speect production model. The most straightforward waveform coding would be to time sample the speech waveform and perform an A/D conversion. The synthesizer would then be a simple D/A converter which could probably be integrated in less than 1000 mils^2 of silicon circuit area. The disadvantage of this approach is that the data rate for reasonable quality speech would be on the order of 30-40 kbits/sec. For voice response applications in which the speech data is stored in ROM this would only allow a limited amount of speech (several seconds) on the largest ROM's available at this time.

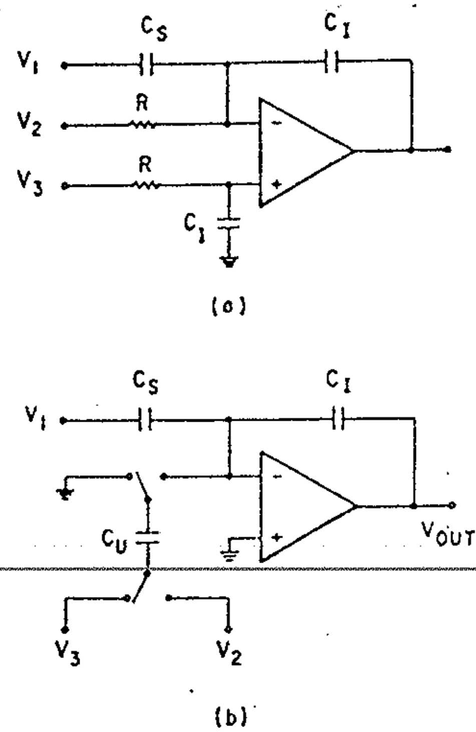

Fig. 2(a) An RC integrator/summer.
2(b) The switched-capacitor implementation.

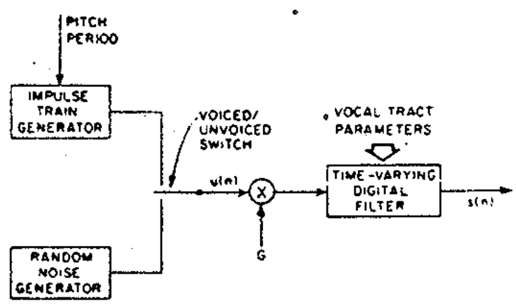

Fig. 3. Block diagram of a basic speech production model.

By assuming a model for the speech production such as shown in Fig. 3, more spohisticated coding techniques can be developed. In this model it is assumed that there is an excitation which can be either a train of periodic pulses or a white noise generator which acts as an input to a time varying filter. The frequency response of this time varying filter contains most of the information about the speech sounds. It fortunately can be quite accurately described by three resonances called formants. The center frequencies of these resonances or formants are an extremely efficient representation of the speech waveform. Using this approach, intelligible speech is possible at rates below 500 bits/sec. [5].

The simplest formant synthesizer is therefore composed of three cascaded second order bandpass sections driven by one of the two possible excitations as shown in Fig. 4.

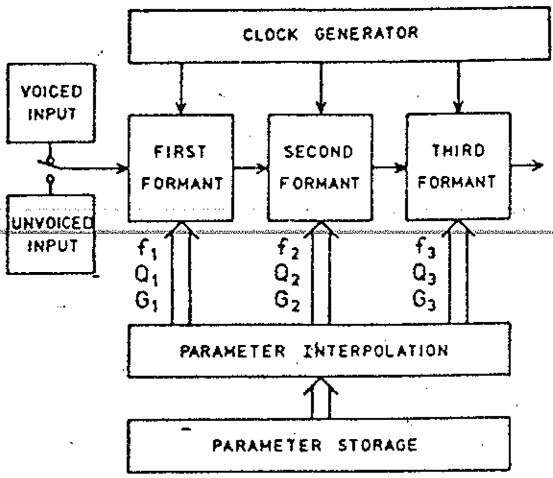

Fig. 4. Block diagram of a simple formant synthesizer.

An all-digital approach to implement the circuit in Fig. 4 is probably possible on a single chip but is very inefficient in comparison to a method which uses programmable switch-capacitor filters as described by D.J. Allstot et al. [6]. Figure 5 is a schematic diagram of one of the programmable formant resonators which has been fully integrated in MOS technology.

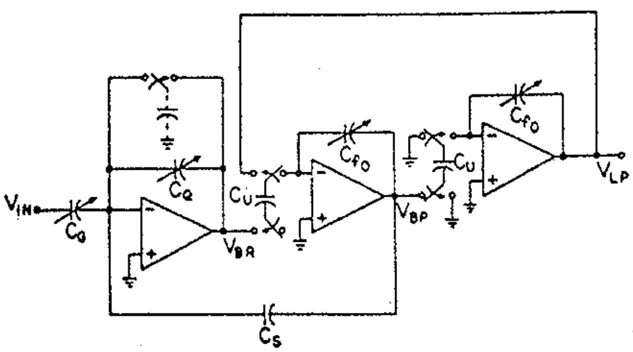

Fig. 5. Schematic of second order programmable switched-capacitor filter.

For the circuit in Fig. 5, the first-order relationships between the programmable capacitor ratios, $\frac{C_Q}{C_s}$ and $\frac{C_G}{C_s}$, and the selectivity (Q) and peak gain (G) are given by

$$\frac{C_Q}{C_s} = Q \tag{5a}$$

and

$$\frac{C_G}{C_s} = G. \tag{5b}$$

In Fig. 6a and b are shown the output of the filter for a variety of values of Q and G.

The first-order relationship between the programmable capacitor ratio, $\frac{C_{f_o}}{C_u}$; the center frequency, f_o; and clock rate, f_c; is

$$\frac{C_{f_o}}{C_u} = \frac{f_c}{2\pi f_o} \tag{6}$$

It was desired to have the center frequencies logarithmically span an octave of frequency such that when the sampling rate was changed by a factor of two, a new set of center frequencies would be obtained which would smoothly continue the logarithmic progression. A series of responses which span an octave are shown in Fig. 6c.

The resonators have their gain, center frequency and selectivity programmed under digital control. In addition, in order to match the long term average power spectra of real speech it was necessary to shape the excitation which was stored in a ROM. The response of the filters are dynamically programmed every 1-20 ms to produce the synthesized speech.

Since the switched-capacitor filters are sampled data filters, the clock frequency can be changed to program the response. This features has been exploited in the formant synthesizer. The first formant stage is either clocked at 10 or 20 kHz, the second at 30 or 60 kHz, and third at 60 or 120 kHz, which results in a two octave range for each formant. Since there are 32768 responses/output/clock frequency, and since there are five different clock frequencies, there are 3.78×10^{22} possible different responses for this system.

As an example of the performance of this system, Fig. 7 shows an English vowel which has been synthesized. It is the vowel /a/ as in "father."

LINEAR PREDICTIVE CODING

Linear prediction of speech (LPC) is an analysis method which extracts the information about the vocal tract transfer function (see Fig. 3) from the speech waveform [7]. Even though

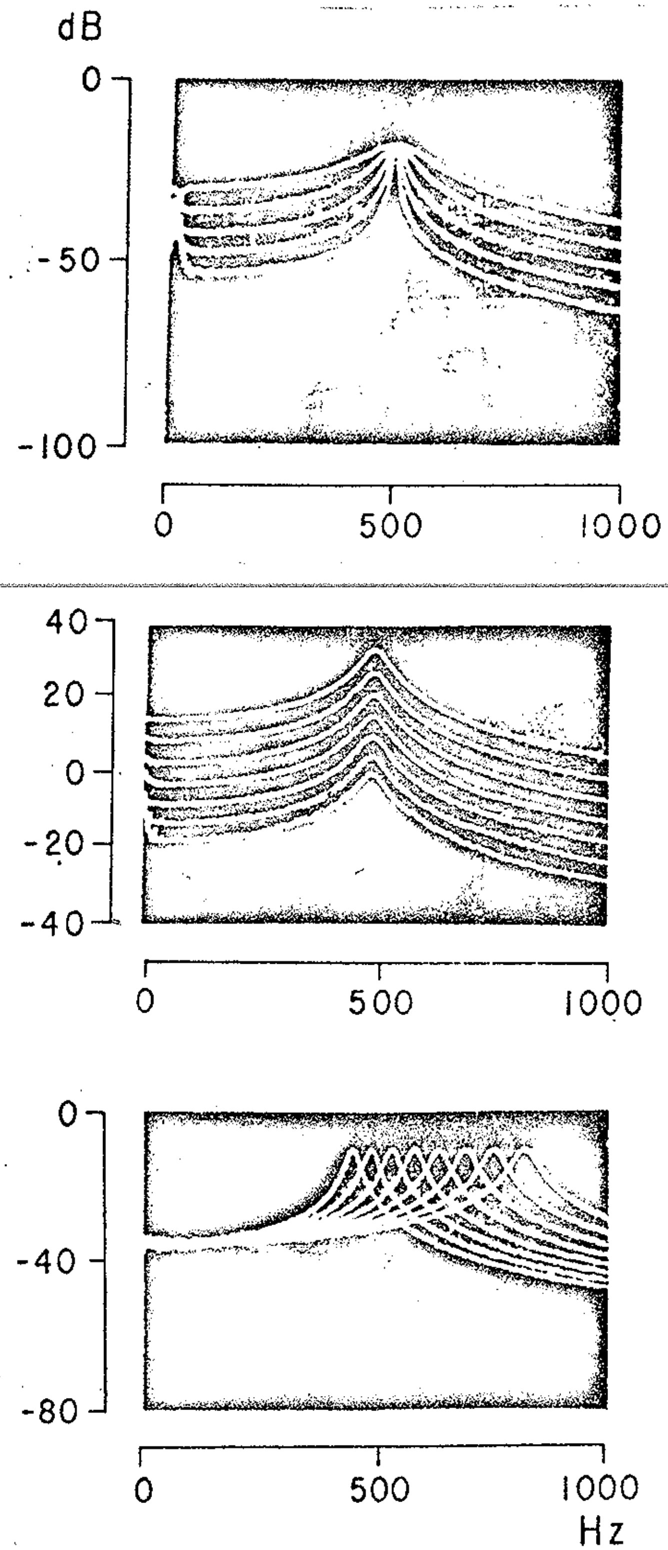

Fig. 6(a). Response of the programmable filter for 5 different values of Q.
6(b) Response for 7 values of gain.
6(c) Response for 8 values of center frequency.

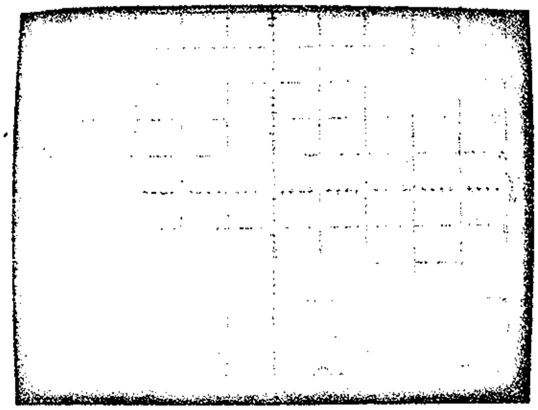

Fig. 7. Spectra of switched-capacitor formant synthesizer for the vowel /a/.

formant center frequencies could be determined in principle from the coefficients obtained in LPC analysis, the extra computation is difficult to perform in real time. The major use of LPC analysis is for very narrow band digital transmission in which highly intelligible speech can be transmitted to a compatible receiver at data rates as low as 2.4 kbits/sec. Another use which is gaining interest is in speech recognition since the LPC coefficients are a very compact representation of the fundamental information of a speech sound.

LPC models the vocal tract resonances by fitting an all-pole transfer function to the vocal tract transfer function. The model has the form (in z-transform notation):

$$H(z) = \frac{1}{1 - \sum_{i=1}^{p} a_i z^{-i}} \quad . \qquad (7)$$

The model parameters (LPC coefficients) are the a_i's. The number of poles used in the model is typically nine to twelve. More poles improve the model accuracy, but a minimum number of poles is desired to minimize the transmission rate.

One approach to the calculation of the LPC coefficients is the autocorrelation approach. This approach requires the computation of p+1 autocorrelation values of the speech waveform computed over a period during which the speech is not changing (i.e. only one sound is being made). The autocorrelation values can be directly transformed into the LPC coefficients by the use of Durbin's (or Levinson's) recursion relation [7]. To accomplish this, the speech is bandlimited to half the sampling frequency and then sampled. A window is then applied to the speech samples to limit the time interval over which the autocorrelation calculation is made. With the window, only 200-300 consecutive speech samples contribute to the autocorrelation function, R(k), where

$$R(k) = \sum_{i=-\infty}^{\infty} \omega(i)s(i)\omega(i+k)s(i+k) \quad k = 0,1,\ldots,p \tag{8}$$

Here $s(i)$ refers to the ith time sample of speech and $\omega(i)$ is the ith sample of a window function. Typically the window is a smooth function which is zero outside of some finite time interval which is usually taken to be approximately 30 ms in length. These functions are called finite impulse response (FIR) windows.

An alternate method for calculating the autocorrelation values was pointed out by Barnwell [8]. He considered the use of an infinite impulse response (IIR) time window. These windows decay to very small values outside of a 30 ms wide region; and thus the results should be very similar to those obtained from the application of a 30 ms FIR window. The advantage of the IIR approach is that the IIR window can be realized by a simple recursive filter which can be efficiently implemented using switched-capacitor techniques.

A brief discussion (following Barnwell) of the form of these filters follows for an IIR time window, $W(z)$. It will be the impulse response of a second order filter having two coincident, real poles:

$$W(z) = \frac{1}{(1-\alpha z^{-1})^2} \tag{9}$$

Its time-reversed impulse response is:

$$\begin{aligned} w(n) &= (1-n)^{-n} & n &\le 0 \\ w(n) &= 0 & n &> 0 \end{aligned} \tag{10}$$

Using equations (8) and (9), we get

$$R(k,j) = \sum_{i=-\infty}^{\infty} s(i)\; w(j-i)\; s(i+k)\; w(j-i-k) \tag{11}$$

Now define

$$s'(i,k) = s(i)\; s(i+k) \tag{12a}$$

$$w'(i,k) = w(i)\; w(i-k). \tag{12b}$$

Then we can write equation (11) as

$$R(k,j) = \sum_{i=-\infty}^{\infty} s'(i,k)\; w'(j-i,k). \tag{13}$$

From the above equation it can be seen that $R(k,j)$ is the convolution of $s'(i,k)$ with $w'(i,k)$. So if we can produce the sequence $s'(i,k)$ and pass it through the linear, time invariant filter with impulse response $w'(i,k)$, we will have calculated the autocorrelation function for lag k (using the data from frame j).

Producing the sequence $s'(i,k)$ requires only multiplication and delay. We need to find the filter corresponding to $w'(i,k)$. Recall that

$$w'(i,k) = w(i)\; w(i+k). \tag{14}$$

Since multiplication in the time domain corresponds to convolution in the z-transform domain, we can get $W_k'(z)$ from equation (14):

$$W_k'(z) = W(z)*z^{-k}W(z) = \frac{1}{2\pi j} \oint W(v)\left(\frac{z}{v}\right)^{-k} W\left(\frac{z}{v}\right)v^{-1}dv. \tag{15}$$

This integral can be evaluated to give

$$W_k'(z) = \frac{(k+1)\alpha^k+(1-k)\alpha^{k+2}z^{-1}}{(1-\alpha^2z^{-1})^3} \tag{16}$$

which yields the transfer function of the filters. Note that each value of k corresponds to a different filter. All the filters have 3 poles at α^2 and one real zero. It has been found experimentally that $\alpha = 0.98$ is the best choice for a 9 pole LPC model with an 8 kHz sampling rate [8].

In order to test the feasibility of this approach, a breadboard version of the complete analysis system has been built. In Fig. 8. a block diagram of the system is shown, in which the sequence of products $s'(i,k)$ defined by eqn. (12a) is produced by multiplication of the input analog signal $s(i)$ and the digitized signal $s(i+k)$ which has been delayed by k samples. The multiplication is performed in a multiplying D to A converter which is time shared to yield the following products: $s(i)\; s(i+k)k = 0,p$ [9]. Each one of these product sequences is multiplexed into a separate switched-capacitor filter with the transfer functions given by eqn. (16).

In Fig. 9 the fit of the transfer function determined by the hardware (the smooth line) is compared to an FFT spectrum of the same speech data. The LPC fit appears smooth because the pitch information which is present in the FFT spectrum has been extracted. Listening tests have been performed on a limited amount of speech which has been analyzed by the hardware and the results indicate that this approach to LPC analysis has the capability of yielding highly intelligible speech.

CONCLUSIONS

The use of switched capacitor techniques have been demonstrated in two applications in speech processing. The ability to perform high performance filtering in the analog domain makes it feasible to fully integrate the two functions (speech analysis and synthesis) onto a small number of integrated circuits. Even though digital techniques could be used throughout, the simplicity and efficiency of using an analog representation, where appropriate, makes it possible to obtain great savings in silicon area. This either allows more functions to be integrated onto a single chip or to reduce the cost of fabrication of the system.

ACKNOWLEDGEMENT

The authors would like to acknowledge the

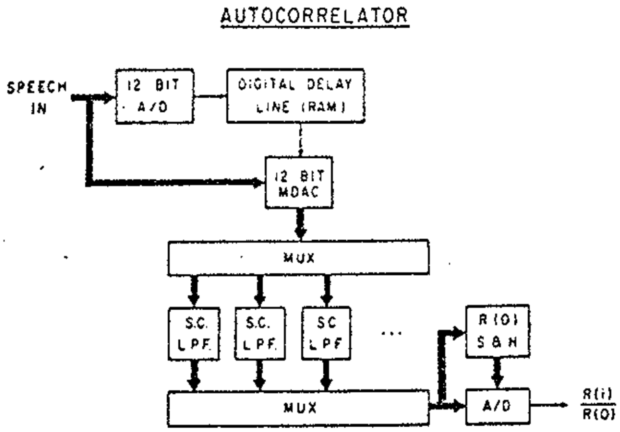

Fig. 8. Block diagram of LPC analysis system using switched-capacitor lowpass filters (SC LPF).

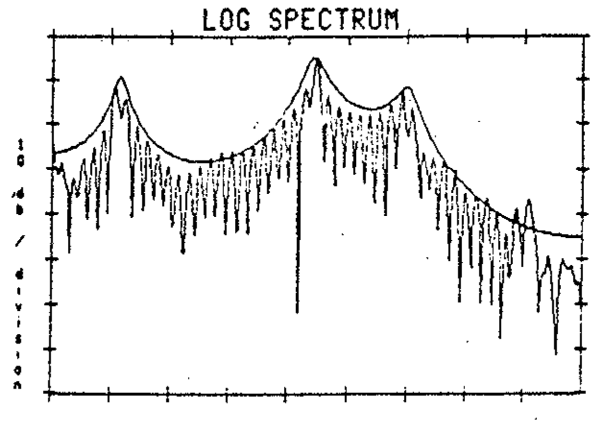

Fig. 9. The fit of the LPC spectra obtained from hardware analysis to a computer generated FFT spectra.

work of E. Evancoe and W. Hasling on the speech synthesizes and of S. Lum and D. Hodges on the LPC analyzer.

This work was supported in part by DARPA under contract MDA903-79-C-0429.

REFERENCES

[1] R. W. Brodersen, P. R. Gray, and D. A. Hodges, 1979, Proc. IEEE, 67, 61-75.

[2] G. M. Jacobs, D. J. Allstot, R. W. Broderson, and P. R. Gray, "Design Techniques for MOS Switched Capacitor Ladder Filters," IEEE Trans. on Circuits and Systems, 1014-1021.

[3] D. J. Allstot and P. R. Gray, "MOS Switched-Capacitor Ladder Filters," IEEE JSSC, 806-814.

[4] B. J. White, G. M. Jacobs, and G. F. Landsburg, 1979, "A Monolithic Dual Tone Multifrequency Receiver," IEEE Int. Solid State Circuits Conf. 36-37.

[5] L. R. Rabiner, and R. W. Schafer, 1978, Digital Processing of Speech Signals, Prentice-Hall Inc., New York.

[6] D. J. Allstot, R. W. Brodersen, and P. R. Gray, 1979, "An Electrically Programmable Analog NMOS-Second-Order Filter," IEEE Int. Solid State Cir. Conf., 76-77.

[7] J. Makhoul, 1975, "Linear Prediction: A Tutorial Review," Proc. of the IEEE, 561-580.

[8] T. P. Barnwell, 1977, "Recursive Autocorrelation Computation for LPC Analysis," Proc. Int'l Conf. on Acoustics, Speech and Signal Processing, 1-4.

[9] S. Lum, 1979, "An Analog-Digital Correlator," Dept of EECS, Univ. of California, Masters of Eng. Project Report.