Design Criteria for New Technologies

Critères de conception pour les technologies nouvelles

A. Maynard Engebretson
Applied Research Laboratory, Department of Computer Science
Washington University

Key words: hearing aid, digital processing, low power VLSI, electronic design

Abstract

In this paper we attempt to identify areas of current hearing aid design that limit performance and that can be improved by applying new technology. First, we will examine hearing aid components, such as microphones, receivers, and batteries. Next, we will address the limits of miniaturization of traditional CMOS technology and discuss the relative merits and problems of analog and digital implementation. Then, we will examine the problems of digital encoding. We will review a number of potentially useful signal processing algorithms related to filtering, feedback stabilization, and noise reduction. Finally, we will compare an analog and digital implementation of a typical hearing aid circuit.

Résumé

Dans le présent document, nous tentons d'identifier les paramètres d'un modèle conceptuel qui limitent le rendement des aides auditives et qui peuvent être améliorés en utilisant une technologie nouvelle. Nous examinons tout d'abord les composantes de la prothèse auditive, comme les microphones, les récepteurs et les piles. Puis, nous parlons des limites de la miniaturisation de la technologie traditionnelle CMOS avant de discuter des avantages et des problèmes reliés aux systèmes analogiques et digitaux. Nous examinons les problèmes de l'encodage digital. Nous nous penchons sur un certain nombre d'algorithmes de traitement de signal pour les aides auditives qui peuvent être utiles pour la filtration, le coulage acoustique et la réduction du bruit. Enfin, nous comparons le comportement électroacoustique d'une aide auditive révélé à l'expertise analogique et digitale.

Introduction

For the purpose of this paper we hypothesize that an ideal hearing aid is one that can be concealed in the ear canal, is efficiently coupled acoustically to the ear, operates on miniscule power, can be programmed with sufficient stable gain to make speech audible, reduces noise present in the signal, and automatically adapts to optimize performance over widely ranging conditions of signal and noise. Whether this ideal hearing aid would provide maximum benefit and satisfaction for the wearer is another issue, which is the focus of a great deal of research and is discussed here and in other papers. The focus of this paper is to address the questions: Can this ideal hearing aid be built; what new technology is available to the hearing aid designer; and what should be the design criteria for this new technology?

The present era of hearing aid development, which began with the invention of the transistor in 1946, is drawing to a close. It is expected that miniaturization of transistor circuitry will reach a point of diminishing return around the middle to end of this decade when feature size begins to approach 0.1 micron. This, of course, will not be the end of further miniaturization for it is likely that new technologies, such as quantum electronics and molecular electronics, will be developed that will result in several orders of magnitude of further size reduction and increased circuit complexity. However, these new technologies are not expected to appear in commercial products for another 10 or 15 years.

There is a great deal left to be done to reduce the size of traditional transistor circuits and improve their performance. Over the next five to ten years we can expect very large scale integration of field-effect transistors to continue to be refined. We can expect design layout tools and simulators to be improved and ASIC (application specific integrated circuit) libraries of functions to be expanded, which will reduce the time and cost of new product development. We can also expect fabrication yields of complex circuits to increase and chip costs to fall. We can expect that high performance, low-voltage, low-power circuits will be developed and that digital processing will become commonplace. The prime mover will probably be high-volume markets, such as personal communicators and audio devices. Hearing aid designs and manufacturing methods are likely to be improved by these new developments. One can expect to see a series of significant improvements in precision, versatility, signal processing complexity, dynamic range, and power consumption. This increased versatility and complexity can lead to "smart" hearing aids that adapt to changing conditions of signal and noise without the need for manual intervention by the wearer.

This paper is organized as follows: First, we briefly examine hearing aid components, such as microphones, receivers, and batteries. Next, we address the limits of miniaturization of traditional CMOS technology. Then, we review a number of potentially useful signal processing algorithms related to

Table 1. Analog amplifier scaling relations (Haskard & May, 1988).

Parameter	Scaling (linear dimension reduced by b)
Area	$1/b^2$ (decreases)
Transconductance	No change
Gain	No change
Gain-Bandwidth Product	b (increases)
Thermal Noise	$b^{1/2}$ (increases)
Flicker Noise	No change
Flicker Noise Bandwidth	No change
Dynamic Range	$1/b^{3/2}$ (decreases)

filtering, feedback stabilization, and noise reduction. Then, we examine the issues of digital encoding. And, finally, we compare the relative merits of an analog and a digital implementation of a hearing aid circuit.

Hearing Aid Components

Microphones

Microphones are almost ideal transducers. They are tiny, wide-band, sensitive and are almost as good as the normal ear (Killion, 1976). There is little to be improved upon with the possible exception of fabricating the microphone on the same chip as the amplifier to reduce overall package size.

Receivers

Receivers are also very good devices. They are reasonably wide-band and small. Exotic magnetic materials have made it possible to reduce the size of the magnetic circuit. The main limitation regarding miniaturization is output power and conversion efficiency, which are related to the acoustic properties of the ear.

Batteries

It is not expected that breakthroughs in battery technology will occur soon. Of possible battery chemistries, the zinc-air battery still provides the greatest energy density currently available. Although the low voltage of zinc-air chemistry is a problem with respect to circuit noise and dynamic range, it is possible to employ efficient on-chip charge pumps that can double and triple and regulate the voltage available on the chip. Therefore, lithium and other battery chemistries, which provide greater voltage but have lower energy densities, are a less attractive solution.

Table 2. Digital inverter scaling relations (Mead & Conway, 1980).

Parameter	Scaling (linear dimension reduced by b)
Area	$1/b^2$ (decreases)
Transit Time	1/b (decreases)
Gate Capacitance	1/b (decreases)
Switching Energy	$1/b^3$ (decreases)

Electronic Circuits

Of the various hearing aid components, the electronic circuitry is likely to undergo the greatest improvement over the next decade. Traditional analog amplifiers lack versatility, programmability, and precision. Discrete-time analog circuits (switched capacitor) suffer from noise and limited dynamic range. These problems will become greater as attempts are made to achieve further miniaturization. As analog and discrete-time circuits are pushed to the limit in terms of complexity, they reach a critical point at which digital processing is a more reasonable approach.

A key part of an analog filter is the operational amplifier. The performance of the op amp determines the limits of performance of the filter. The effects of reducing the size of the amplifier circuit can be estimated based on certain scaling relations shown in Table 1 (Haskard & May, 1988). From these relations it can be seen that the overall effect of reducing the size of analog circuits is to increase noise and decrease dynamic range. Dynamic range decreases as a 3/2 power of the linear scale factor (9 dB for each reduction of linear dimensions by 2). Because dynamic range is already marginal in hearing aid circuits, it seems unlikely that significant size reduction, translating into increased complexity of function, is possible. Other factors in addition to dynamic range enter into scaling of analog circuits. For example, component matching is more difficult to achieve as one approaches the limits of line definition of the fabrication process.

Scaling considerations for digital circuits are more encouraging. The dynamic range of a digital amplifier (inverter) need only be great enough to resolve two levels. As the size is decreased, more bits can be added to increase the dynamic range of the signal representation. As the linear dimensions of digital circuits are reduced, propagation delay and gate capacitance are reduced proportionately as shown in Table 2 (Mead & Conway, 1980). Furthermore, the switching energy is reduced by the third power of the size. Therefore, an inverter fabricated in a 1 μm process will require one-quarter the area and one-eighth the switching energy of one fabricated in a 2 μm process. Because the switching energy scales by a power of three and the area by a

power of two, it turns out that the complexity of the circuit can be increased with an overall savings in power consumption by making it smaller.

For the purpose of estimating performance of digital processing structures we will assume that a digital inverter (two transistors) occupies an area of 15λ by 15λ on the average and that its switching energy is on the order of $5x10^{-15}$ λ^2 joules, where λ is expressed in μm (Mead & Conway, 1980). These are average values that can give us a crude estimate. Using these numbers, 4000 inverters switching at a 1 Mhz rate will consume about 540 μw of power if they are fabricated using a 3 μm process. The 4000 inverters will occupy an area of about 14 mm^2. At 0.5 μm these 4000 inverters will consume about 2.5 μw and occupy an area of about 0.4 mm^2. Therefore, power consumption is really not the problem as it was once thought to be because, by using small feature size and low voltage, it is possible to design useful digital processing structures that consume miniscule amounts of power. We shall see in a later section (Comparison of Analog and Digital Implementations) an example of a digital processing structure that can be designed to be more versatile and require less area and power than its analog counterpart.

Signal Processing

Although digital signal processing structures can be implemented in low-power form, this new technology has not yet been exploited in commercial hearing aids. This is surprising because digital circuits are relatively easy to use and offer a number of advantages with regard to functionality and precision. Analog-to-digital and digital-to-analog converters are still a problem, but solutions are near at hand (Engel, 1990). Typical types of converters are discussed later (Encoder and Decoder Considerations). Typical design criteria for digital filters are discussed below.

Digital filters can be implemented as fixed or adaptive structures. Examples of fixed filters are infinite-impulse-response (IIR) filters [recursive filters] and finite-impulse-response (FIR) filters [non-recursive]. Recursive filter designs generally require fewer arithmetic operations and yield a sharper amplitude response characteristic whereas non-recursive filters offer the advantage that they are inherently stable so that coefficient accuracy is less important. If linear phase is important, a symmetrical form of an FIR filter is also best. Note that linear phase characteristics are desirable with multiple-channel hearing aids in order to avoid aberrant interactions between channels at band edges. IIR filters are useful for implementing impulse-invariant designs having analog counterparts. Examples include a circuit that measures signal energy (a rectifier followed by a first-order low-pass filter) and various other types of analog filters (such as band-pass, high-pass, low-pass, Elliptic, Tschebycheff & Butterworth) that are familiar to designers. FIR filters are useful for implementing filters that provide gradual shaping of the frequency response to achieve a desired prescriptive insertion frequency-gain function.

Typical applications of adaptive filters include: (1) equalizing filters that adapt to changes in the external feedback path from receiver to microphone, and (2) inverse linear predictor filters that adapt to peaks in the noise spectrum and reduce the gain of the hearing aid at those frequencies. An example of an adaptive method that is well suited for hearing aid applications is the least-mean-square (LMS) algorithm (Widrow et al., 1975) which can be simplified to a binary form requiring no multiplications to update the coefficients.

Design Criteria for Fixed Filters

The frequency resolution of an FIR filter is related to the length (number of taps) of the filter. Two frequencies of equal amplitude separated by Δf can be resolved by a filter with $N=f_s/\Delta f$ taps where f_s is the sampling frequency. If the relative amplitudes of the two frequencies are greatly different, additional filter taps are required to resolve the two signals because of the gradual slope of the filter skirts.

One simple way to compute the desired coefficients of an FIR filter, known as the window method, is based on the fact that the impulse response of an FIR filter is the coefficient sequence itself. One simply determines the impulse response of the desired filter response at uniformly spaced frequency intervals of $1/f_s$. Because the desired impulse response will be infinitely long, it must be truncated to create a finite sequence of coefficients for the FIR filter. A window function is generally used to avoid the discontinuities of truncation. The window function increases the steepness of the skirt at band edges and also reduces the frequency resolution of the filter. A Kaiser-Bessel window ($\alpha=2.5$) provides a good compromise between frequency resolution and filter skirt steepness. It has a resolution of about 2.2 bins and an out-of-band rejection of about -57 dB, where a bin width is defined as f_s/N, N is the total number of filter taps, and fs is the sampling frequency (Harris, 1978).

An expression that was derived by Kaiser (1974), based on the window design method, for relating filter characteristics to number of filter taps is:

$$N=\frac{-R-7.95}{14.36\,\Delta f/f_s}+1$$

where R can be thought of as the desired out-of-band rejection in decibels, Δf is the desired transition band, f_s is the

sampling frequency, and N is the required number of taps. As an example, a filter with 31 taps operating at a sampling rate of 16 kHz will have an out-of-band rejection of -40 dB and a transition band of about 1.2 kHz. Details of the window design method are given in Rabiner and Schafer (1978).

As stated above, an advantage of FIR filters is that the frequency response can be shaped to compensate for the frequency-dependent loss of sensitivity of the ear. An example is the CID digital hearing aid (Engebretson, Morley, & Popelka, 1987), which utilizes four channels of 31-tap, symmetric, FIR filters. The desired prescription for insertion gain for a patient is specified at audiometric frequencies. The corresponding filter response is then derived by a PC-based fitting system from this specification, and the time sequence of the desired impulse response is derived by computing an inverse DFT. The resulting impulse response is windowed to obtain the FIR filter coefficients, which are then down-loaded to the hearing aid. By using this method, the desired overall insertion gain of the digital hearing aid can be achieved to within a few decibels of rms error (French-St. George & Metzger, 1990).

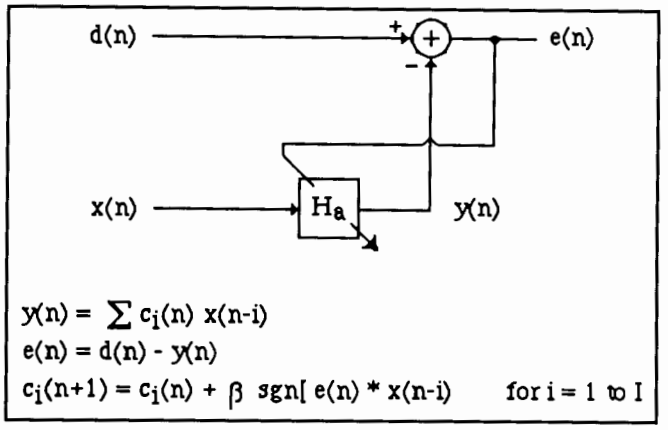
One problem with the FIR filter is that a large number of taps are required to create a narrow-band filter. For example, a one-third octave band centered at 400 Hz is about 90 Hz wide. To create an FIR filter with a 90 Hz band-width requires about 300 taps at a sampling rate of 16 kHz, which is impractical to implement in an ear-level hearing aid circuit. Instead, IIR filters are potentially better.

The design of IIR filters is straightforward (e.g., see Daniels, 1974). Typically, one starts with an analog filter design that has the desired characteristics. Then the analog design is converted to digital form using one of many types of analog to digital transformations. The recursive digital expression for the IIR filter is derived directly from the digital transform. An example of a recursive filter is a simple third-order, low-pass, Butterworth filter, which is used later in this paper for comparing analog and digital implementations. The recursive expression for the digital filter is:

$$y(n) = a\,x(n) + b\,x(n-1) + c\,y(n-1) + d\,y(n-2) + e\,y(n-3)$$

which requires 5 multiplications, 4 additions, and 5 storage registers. This is a computational load equivalent to about 5 taps of an FIR filter. One problem with IIR filters is that when narrow bands or steeply sloping skirts are required, the coefficients must be specified with a high degree of precision, which is difficult to achieve in practice. If the coefficients are specified to a lower precision, the filters may be unstable or may generate excessive arithmetic noise. The precision requirement is less severe for wider band filters with more gradual sloping skirts.

Figure 1. Basic adaptive LMS algorithm. Coefficients of the adaptive filter, Ha, change in the direction to minimize the mean-squared difference between the desired response, d(n), and the filter output, y(n). Coefficient updating is accomplished by taking the exclusive-or of the sign bits of e(n) and x(n-i).



Design Criteria for Adaptive Filters

A number of adaptive filter algorithms have been studied by various designers. One algorithm that appears to be practical to use in a hearing aid is a simplified version of the LMS algorithm (Widrow et al., 1975) that is illustrated in Figure 1. The FIR filter in this figure has an output:

$$y(n) = \sum c_i\, x(n-i)$$

which is a filtered version of the reference signal, x(n). The sequence, d(n), is the desired signal that the adaptive filter is trying to achieve. The error between the filter output and the desired signal is:

$$e(n) = d(n) - y(n).$$

The mean of the squared error, $e^2(n)$, forms a dish-shaped function in coefficient space that has a unique minimum. The coefficients can be adaptively driven towards this minimum by using the gradient of the error to determine the direction of steepest descent. A simplified algorithm for adjusting the coefficients that requires no multiplications is given by:

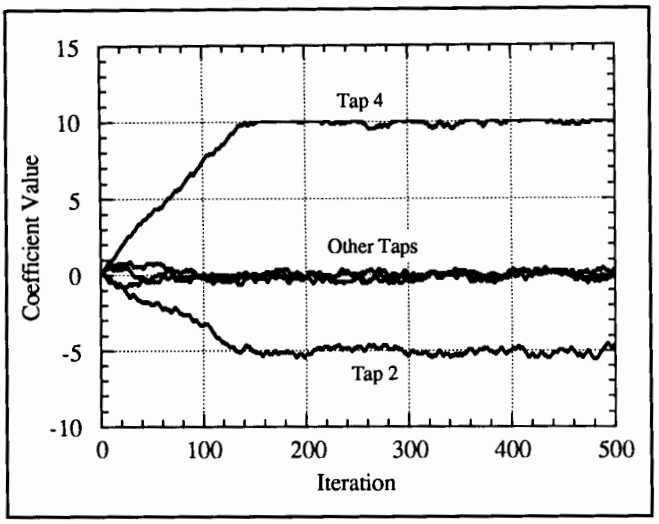
$$c_k(n+1) = c_k(n) - \beta\ \mathrm{sgn}[\,x(n-k)\,e(n)]$$

where β can be a fixed power of two. This is equivalent to incrementing or decrementing the coefficient register depending on the value of the sgn function. The sgn function is a simple exclusive-or of the sign bits of e(n) and x(n-k).

The behaviour of the LMS algorithm is illustrated in Figure 2. For this illustration the desired signal, u(n), is derived from the model expression:

$$u(n) = 10\,x\,(n-2) - 5\,x(n-4)$$

Figure 2. Typical simulation results for the model, y(n) = 10 x(n-4) - 5 x(n-2). Coefficients follow a linear path during adaptation until they converge to the model values. Once the model values are reached, further adaptation results in a dithering of coefficient values around the correct values.

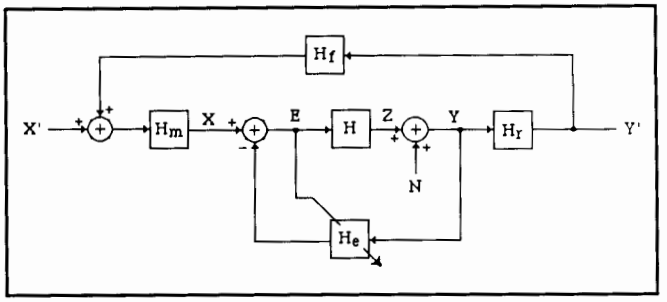


where x(n) is the random sequence also serving as the input to the adaptive filter. It can be seen that the coefficients follow essentially a linear path in adapting to the model and that once the correct values are reached, the algorithm randomly dithers around the least significant bits of the coefficients. The sgn function simplification of the LMS algorithm results in a slow rate of adaptation. However, this is desirable in many applications. The simplified algorithm has other desirable characteristics. For example, no dead zone occurs as the error becomes exceedingly small to cause a coefficient tracking offset. The adaptive filter algorithm is robust and versatile and can be used for feedback equalization and noise reduction as is described below.

Adaptive Feedback Equalization

Feedback instability is a persistent problem with hearing aids, and active cancellation of the feedback path has been proposed as a solution (Egolf, 1984; 1986). The feedback path is generally unknown and is continually changing, and an effective equalization filter must arrive at a correct estimate of the feedback path and adapt to changes in the path characteristic. The algorithm described above can be used for adaptive feedback equalization as shown in Figure 3. The reference signal, x(n), in Figure 1, is derived from the output signal of the hearing aid, Y, in Figure 3, and the desired signal, d(n), in Figure 1, is the signal, X, in Figure 3, which includes the external feedback signal that is correlated with the output signal of the hearing aid, Y. The output also contains a pseudo-

Figure 3. Functional block diagram of a hearing aid system with adaptive feedback equalization. Primed symbols represent acoustic variables. Unprimed symbols represent digital values. Hm and Hr represent the transfer characteristic of the microphone and receiver, respectively. Hf and He represent the feedback path and the equalization filter, respectively. The desired acoustic frequency response of the hearing aid is determined by filter, H, and is the product, Hm*H*Hr.

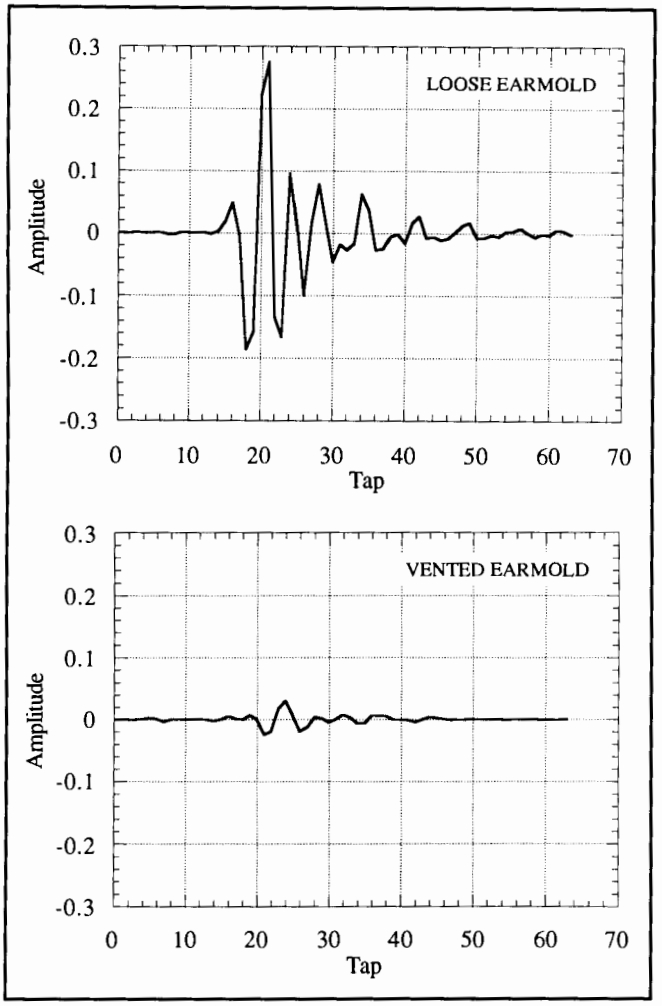


random noise sequence, N, that is uncorrelated with all other signals in the system and that is set to a level near the listener's threshold to excite the adaptive process when no input signals are present or when they are at a low level. The signal, X, also contains externally derived signals that are delayed by the main hearing aid filters and are, therefore, relatively uncorrelated with Y. Because of the slow adaption rates that are used (several seconds), only periodic external signals that persist will upset the equalized state of the system.

In studies using the feedback equalization algorithm with hearing impaired subjects, it was found that stable gain margins of hearing aids can be improved by 10 to 15 dB with adaptive equalization (Engebretson et al., 1990; 1991). This has also been the finding of Dyrlund and Bisgaard (1991). Not only does the algorithm suppress oscillation, but it also equalizes the under-damped, non-oscillatory feedback condition that often occurs at high gains and tends to degrade hearing aid performance. A drawback of the algorithm is that it will also cancel certain sounds that have a long-term stationary characteristic. Some alarm sounds may fall into this category.

With the CID digital hearing aid it is possible to up-load the equalization filter coefficients to a host computer. Figure 4 illustrates the impulse response of the equalization filter for two conditions of acoustic leakage with a KEMAR manikin test setup after the equalization filter has reached a steady state. The impulse response represents the external feedback characteristic. Most of the delay before the start of the impulse response is due to the receiver (each tap corresponds to a 60 μs delay). The equalization filter has to be long enough to encompass the impulse response of the acoustic feedback path. In our experience, this is on the order of 50 or 60 taps for a behind-the-ear hearing aid.

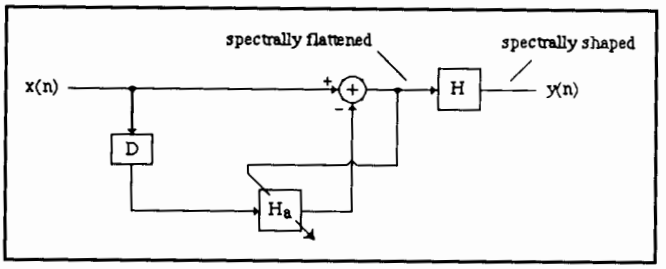
Figure 4. Typical results of adaptive feedback equalization using a digital hearing aid and KEMAR test setup. Curves shown are up-loaded coefficient values of the equalization filter after having converged to a stable state for conditions of (a) a loose earmold and (b) a tight-fitting, vented earmold.



Adaptive Noise Reduction

Another commonly cited problem with hearing aids is background noise. Limitations of methods of noise reduction are discussed in the National Research Council (1989) report and have been reviewed by Lim and Oppenheim (1979). There appears to be no effective way to eliminate noise once it has corrupted the signal without having access to a good reference signal of the noise. Unfortunately a good noise reference is not available in hearing aid applications. Therefore, the only practical means for reducing noise with a hearing aid seems to be one of filtering the corrupted signal at frequencies in which the noise energy is substantial. Figure 5 illustrates how the adaptive algorithm described in Figure 1 can be used for noise reduction. In this case the reference signal, x(n), in Figure 1, is the same as the desired signal, d(n), in Figure 1,

Figure 5. Functional block diagram of adaptive noise reduction algorithm. D represents a one-sample delay, and Ha represents the adaptive filter. The algorithm tends to flatten the spectrum of the input signal, x(n). H is a fixed, programmable filter for shaping the flattened spectrum to satisfy the conditions of audibility for the hearing impaired listener.



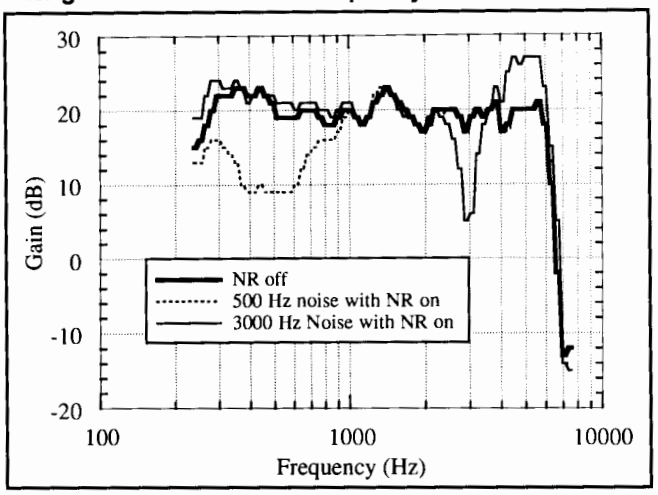
except that it is delayed by one sample. The adaptive filter of Figure 5 then becomes an inverse linear prediction filter. By using a slow rate of adaption the filter adapts to the long-time average spectrum of the input signal. The spectrum at the output of the inverse filter is thereby flattened and can be shaped by a second filter to satisfy the prescriptive needs of the patient.

When noise is introduced at the input to the system, the inverse filter will adjust itself to reduce the noise energy. The signal energy will also be reduced at the same frequencies as the noise. However, the overall signal-to-noise ratio of the corrupted signal is improved if the noise energy is concentrated in a narrow band of frequencies. This is illustrated in Figure 6, which shows the filter response to one-third octave bands of noise centered at 500 Hz and 3000 Hz. By comparing the two curves (without noise and with noise at 500 and 3000 Hz) it can be seen that only the frequencies in which noise is present are filtered, and the higher frequencies are essentially unchanged.

Most noises have energy that is concentrated in the low frequencies. Examples are fan noise, motor noise, and highway noise. For these types of noise the algorithm acts as a high-pass filter with an adaptive low frequency characteristic. Other types of noise, such as speech babble, have greater band width. However, the long-time average spectrum of speech babble tends to decrease with frequency at a greater rate than the speech of a single talker and, in addition, room reverberation tends to increase the noise level at lower frequencies. Therefore, speech babble can be considered to be low-pass noise.

Although it has not been demonstrated conclusively, this form of adaptive noise reduction has the potential for providing benefit for the hearing impaired listener in noisy situations. We have found that it improves intelligibility by increasing the gain of the system (audibility) in the mid to high range of

Figure 6. Examples of adaptive filtering of system in Figure 5 for signal alone (heavy solid curve) and for signal mixed with one-third octave band of noise centered at 500 Hz (dotted curve) and 3000 Hz (thin solid curve). Note that filter gain is reduced at the frequency of the noise bands.



frequencies (French-St. George, Engebretson, & O'Connell, 1992). Because the adaptive filter flattens the spectrum, it increases the consonant-to-vowel ratio, which may also improve speech intelligibility. The algorithm also reduces low-frequency energy, which is thought by some to mask high-frequency speech sounds through upward spread of masking (Rankovic, Freyman, & Zurek, 1992). In addition, the algorithm reduces the level of noise presented to the listener as compared with that of a linear amplifier. This should provide greater comfort and reduce listener fatigue.

Encoder and Decoder Considerations

A key element of any digital processing system is the circuit that converts the signal from analog to digital form and from digital to analog form. In a digital hearing aid we desire a converter with a wide dynamic range that draws small amounts of power and can be fabricated monolithically in a small size. A number of design approaches are possible. Engel (1990) has implemented a digital-to-analog converter that utilizes a charge redistribution method that satisfies the power constraints of a hearing aid and, in addition, yields a logarithmic representation that has certain other advantages with regard to signal processing. Engel's approach is to use a large capacitor that is charged to a reference voltage and a smaller capacitor that repeatedly discharges the larger one in small, constant ratio nibbles. The resulting discharge curve of voltage across the larger capacitor follows an exponential curve that is determined by the ratio of the small to the large capacitance. The final discharged state of the large capacitor is determined by the number of nibbles taken, which is controlled by a counter that is initialized with the log digital value of the signal. The log base is the ratio of the capacitance of the large capacitor to the sum of the capacitance of the large and small capacitors. Because monolithic capacitors can be matched to within 1/2%, an accurate log base is relatively easy to achieve.

Another form of charge redistribution decoder is one that uses a network of binary-weighted capacitors. Each capacitor is charged to a reference voltage or discharged to ground, depending on the bit value that is represented by the capacitor. Then all the capacitors are connected in parallel. The resulting voltage across the parallel combination is the analog output signal. Because capacitor values on the same chip can only be matched to within 1/2%, this method is limited to between 40 and 50 dB of dynamic range (8 or 9 bit resolution).

Analog-to-digital converters (ADC) are generally implemented by adding a comparator and control logic to the DAC circuitry. The control logic systematically sequences the DAC until it's output is the same as the unknown input signal. A variety of encoder methods are used. The most common is the trial and error method (successive approximation) that requires a number of trials equal to the word length of the encoder.

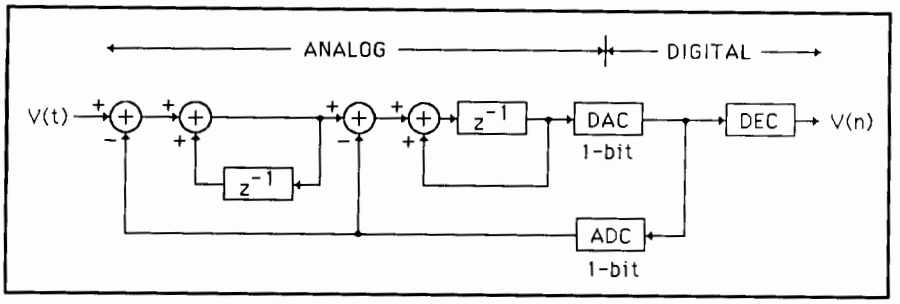
With regard to encoders and decoders, circuit elements, such as capacitors and transistor switches, are not ideal elements because of parasitic capacitance, switch resistance, voltage offset, and noise. Therefore, charge redistribution schemes are generally limited to 8 or 9 bits (about 50 dB dynamic range). However, a greater dynamic range is desired of a hearing aid circuit. In fact, it can be argued that a greater dynamic range and a lower noise floor is required for hearing impaired listeners than for normal hearing listeners because of the greater disparity of sensitivity of the impaired ear at different frequencies and because of the differences in masking functions of the impaired and normal cochleas.

A relatively new approach to the ADC problem that appears to be potentially useful for hearing aids is the delta-sigma encoder. Delta-sigma encoders have revolutionized audio recording and have made possible high-performance CD and DAT audio systems. An advantage of delta-sigma is that only one circuit element is critical in fabrication. However, it must oversample the input signal at high clock frequencies.

The general approach to an over-sampling delta-sigma encoder is illustrated in Figure 7. The feedback circuits on the analog side are simple discrete-time filters that shape the quantizing noise so that most of the noise energy is moved to higher frequencies that lie above the signal base-band. The quantizing noise is the difference between the analog input and quantized output, which in this case has only two states (1 bit). The sigma-delta encoder is followed by a low-pass

decimation filter that removes the high-frequency quantizing noise, extends the word-length, and reduces the sampling rate. A variety of clever, simple digital circuits have been used for the decimation filter. A second-order delta-sigma encoder operating at an over-sampling ratio of 64 has a signal-to-noise ratio of about 80 dB (Candy & Temes, 1992). Assuming a base-band sampling rate of 20 kHz, the over-sampling frequency is 1.28 MHz, which should be reasonable to achieve in a low-power hearing aid form.

Figure 7. Functional block diagram of a delta-sigma encoder. The spectrum level of the quantizing noise of the one-bit converter is reduced by using a sampling rate much higher than the signal band-width. The quantizing noise in the signal band is further reduced with the simple analog filter by shaping the noise so that it is concentrated outside the signal band. The decimation filter is a simple multi-rate, low-pass digital filter that removes the high-frequency noise energy.



An important design parameter of converters is their dynamic range, which is often taken as the ratio of the rms value of the largest sinusoidal signal that can be encoded without peak clipping to the rms quantizing noise of the encoder. The dynamic range of an encoder can be estimated by the expression:

$$DR\ (dB) = 6\ (N-1) + 11 - 3$$

where N is the number of bits used to represent the converted value. The factor of 11 in the expression accounts for the statistical distribution of the quantizing error and the factor 3 is the peak-to-rms ratio for a sinusoid. Therefore, a 12-bit encoder will have a dynamic range of about 74 dB. Assuming that the encoder can accommodate a maximum sinusoidal signal of 1v rms, the quantizing noise will be -74 dBv or 200 μv rms. This quantizing noise has a uniform spectrum so that the spectrum level of the noise, QN_{SL}, can be computed from the expression:

$$QN_{SL}(dBv) = MAX(dBv) - DR(dB) - 10\log_{10}(f_s/2)$$

where MAX is the maximum rms sinusoidal signal level that can be accommodated by the encoder without peak clipping, DR is the dynamic range as given above, and f_s is the sampling frequency of the encoder. The perception of the noise and its masking effect is probably more related to the noise energy in critical bands of the ear. Therefore, the noise in one-third octave bands is probably a more relevant design criteria. The one-third octave band quantizing noise is given by the expression:

$$QN_{1/3}(dB) = QN_{SL}(dBv) + 10\log_{10}(f) - 6.35$$

where f is the center frequency of the one-third octave band of measurement and the factor, 6.35, takes into account the difference in energy in an octave band and a one-third octave band centered at f.

The above expressions are for the case of linear encoding where the signal values are represented in linear form. In the linear case the signal-to-noise ratio, SNR, of the encoder is a function of the signal level and is given by:

$$SNR(dB) = SIG(dBv) - MAX(dBv) + DR$$

where MAX and DR are as defined above, and SIG is the rms signal level. The situation is different when the signal values are represented in logarithmic form. In the logarithmic case the SNR is the same as for the linear case at small signal levels, but at larger signal levels the SNR reaches a maximum. SNR functions of input level are shown in Figure 8 for a 12-bit linear encoder and a 9-bit log encoder. The maximum SNR of the logarithmic encoder is given by:

$$SNR_{max}(dB) = 11 - 20\log_{10}[|\ \ln(B)\ |\ ]$$

where B is the log base. The factor of 11 is again due to the statistical distribution of the noise energy. One advantage of using logarithms is that multiplication is then a simple add operation, which greatly simplifies the digital filter circuitry. Another advantage is that the dynamic range can be large. The dynamic range is given by:

$$DR_{log}(dB) = 2^{N-1}\ [20\log_{10}(B)]$$

where B is the log base and N is the number of bits representing the magnitude of the log value. For a log base of 0.94 and a 9-bit log representation (8-bit magnitude plus sign), the maximum signal-to-noise ratio is 35 dB and the dynamic range is 135 dB!

A comparison of the three encoder methods discussed above is shown in Figure 9. The encoders have been scaled so that the maximum signal that can be accommodated without clipping corresponds to 120 dB. The signal level in this figure corresponds to 100 dB SPL. The one-third octave band noise of the logarithmic encoder is highest. However, it should be noted that the noise of the logarithmic encoder is a function of signal level and, therefore, will be less for a smaller signal. The noise of a 12-bit linear encoder is lower than that of the log encoder, and the noise of the second-order delta-sigma encoder with an over-sampling ratio of 64 is intermediate.

Figure 8. Comparison of the signal-to-noise ratio (SNR) as a function of signal level for a linear analog-to-digital converter (ADC) and a logarithmic ADC. The maximum SNR that can be achieved with logarithmic encoding is a function of the log base.

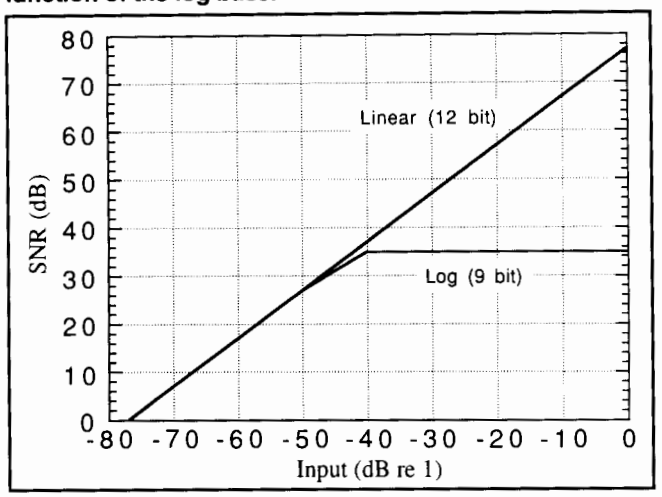
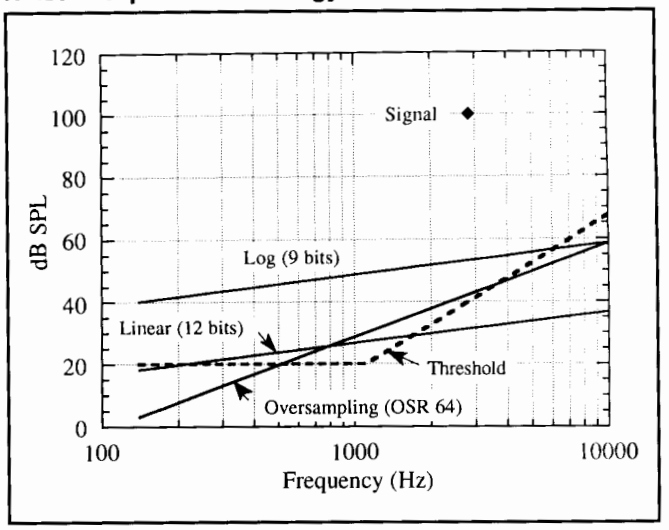


Figure 9. Quantizing noise resulting from a 100 dB SPL sinusoid for linear, logarithmic, and delta-sigma encoders compared with a typical sloping hearing loss threshold. Noise is expressed as energy in one-third octave bands.



Comparison of Analog and Digital Implementations

In this section we compare analog and digital methods of implementation of a programmable hearing aid circuit. The circuit is a programmable low-pass filter in series with a programmable high-pass filter. The analog version is a switched-capacitor implementation that was developed in the mid 1980s and is described in Callias et al. (1989). The cut-off frequency of the high-pass and low-pass filters are programmable in eight steps from 1 to 8 kHz and 0.25 to 4 kHz, respectively. The filters are third-order Butterworth types. The chip containing the filters is reported to operate over a power supply range of 2.4 to 3.1 volts and consume 200 μA of current. The total chip area is 9.3 mm^2. The chip was fabricated in a 3 μm CMOS process. The dynamic range of the filters is stated to be about 65 dB. The 3 v power for the filter circuit is obtained from a 1.2 v hearing aid battery by the use of a voltage tripler-regulator circuit that resides on another chip of the set. A preamplifier that resides on a third chip has an input equivalent noise less that 3 μv over a band-width of 10 kHz. When connected together, the entire chip set has an equivalent input noise of about 4 μv and consumes about 1.5 mA from 1.3 v battery power.

How does this compare with a digital implementation? As has been described above, each of the third-order Butterworth filters can be implemented by the recursive expression:

$$y(n) = a\,x(n) + b\,x(n-1) + c\,y(n-1) + d\,y(n-2) + e\,y(n-3).$$

It can be seen that five multiply-add operations are required per filter per sample for a total of ten operations per sample. To achieve a band-width of 8 kHz requires a sampling frequency of greater that 16 kHz. The following design projections will assume a sampling frequency of 20 kHz, which corresponds to a sample interval of 50 μsec. Therefore, each of the ten multiply-add operations must be performed within 5 μsec.

An interesting design for the digital filter is a bit-serial structure illustrated in Figure 10. A bit-serial implementation might be appropriate because the processing complexity of the filters is modest, and a bit-serial implementation results in a small size. SR3 in Figure 10 is a bit-serial shift register that contains two consecutive input samples; SR1 is a shift register that contains the partial sum of products of the recursive expression; and SR2 is a shift register that contains the last three output samples. The EEROM is an electrically-programmable, nonvolatile, read-only memory that contains the coefficients that are down-loaded from the fitting system. The multiplexors (MPX's) are simple one-bit switches that control the flow of bit streams. The controller logic, which can be a finite-state-machine implemented as a programmed logic array structure, is not shown in Figure 10.

The design criteria that determine the length of the registers are as follows: The length of the input register (SR3) is determined by the desired dynamic range of the encoder and filter. If we assume an 80 dB or better input dynamic range, the samples must be at least 14 bits. Therefore, two input samples, x(n) and x(n-1), require an input shift register that is 28 bits long. The same dynamic range is required for the output register (SR2), and because three delayed values of output, y(n), y(n-1), and y(n-2), are required for the recursive expression, the output register must be 42 bits long. In order

Figure 10. Functional logic diagram of bit-serial, impulse-invariant implementation of a third-order Butterworth digital filter.

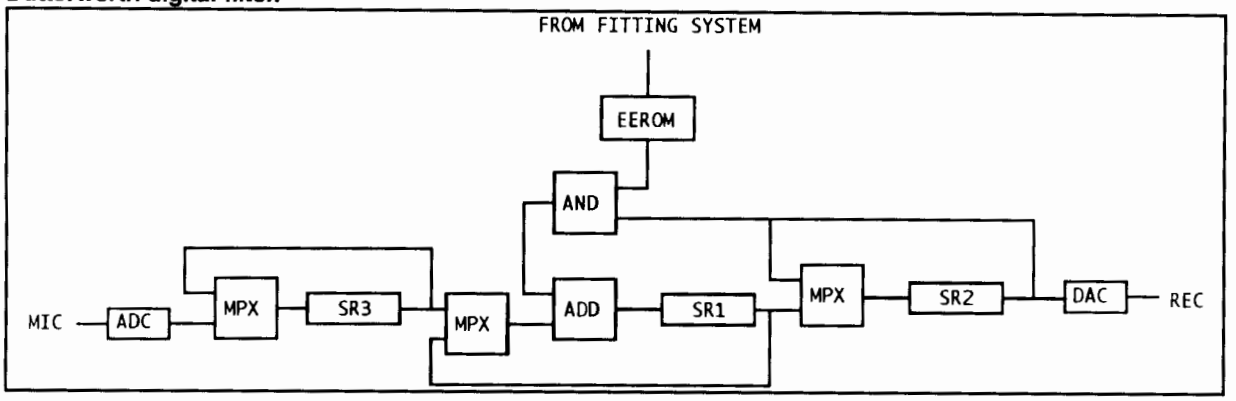


Table 3. Summary of impulse-invariant filter design (number of devices).

Device	Number of stages	Transistors per stage	Total number of transistors	Area $l = 1\ \lambda m$
SR1	30	14	520	
SR2	42	14	588	
SR3	28	14	392	
ADDER	1	40	40	
Total			1540	0.35 μm2

Table 4. Summary of impulse-invariant filter design (state transitions and power).

Device	Shifts per sample period	Transistor state changes per sample period	Switching power $\lambda = 1\ \mu m$ fs = 20 kHz
SR1	30x14x5-2100	2100x520-1,092,000	
SR2	42	42x588-24,696	
SR3	28	28x392=10,976	
ADDER	2100	40x2100=84,000	
Totals		1,211,672	121 μW

to avoid round-off errors of partial sums of the multiplier/accumulator, the accumulator register (SR1) should probably be at least as long as a product term, with additional guard bits to avoid overflow of the partial sum. Assuming 14-bit samples, 14-bit coefficients, and 6 guard bits yields a 30-bit accumulator register length.

With regard to timing, register SR3 is shifted 28 bit positions during each sample period, the output register is shifted 42 bits, and the accumulator register, SR1, is shifted 30 bits for each of the 14 coefficient bits for each of the 5 coefficients for a total of 2100 shifts per sample period. This is summarized in Tables 3 and 4.

The power dissipation of a CMOS implementation is related to the switching energy of the transistors used in implementing the various logical functions and the rate at which the transistors are turned on and off. A serial shift register stage can be implemented with 14 transistors, a single-bit adder and carry register with 40 transistors, and a multiplexor with 4 transistors. The adder changes state at the rate of 2100 changes per sample. Therefore, the total number of transistor state changes per sample is about 1.2 million. Using our estimate cited previously and assuming a sampling rate of 20 kHz, this corresponds to a total switching power of about 121 μw with a feature size of 1 μm. This would be about 968 μw with a feature size of 2 μm. The size of the structure can be roughly estimated by assuming that a transistor can be laid out with a density of about 15x15 λ^2 square. Because the total number of transistors is 1,540, the total area can be estimated to be 0.35 mm^2 for a feature size of 1 μm. The area increases to 1.2 mm^2 for a feature size of 2 μm.

This simple digital processing structure can simulate the characteristics of the switched-capacitor analog filter described above as accurately as desired. The high-pass and low-pass cutoff frequencies of the two filters are determined by the coefficients of the recursive expression that are down-loaded to the EEROM memory. Furthermore, because the recursive expression encompasses a wider selection of filter types, other filter types can be chosen by down-loading appropriate values for the coefficients.

Another approach, which is not an exact simulation of the Callius et al. hearing aid circuit, but which accomplishes the same end, is the use of an FIR filter, illustrated in Figure 11. The goal here is to try to duplicate the characteristics of the Butterworth filters with an FIR filter. A third-order Butterworth filter has a characteristic that is flat within 3 dB from 0 to 1 kHz, falls to 18 dB at 2 kHz, and continues to fall off at 18 dB/octave for frequencies above 2 kHz. Using the window design method described previously, a symmetric FIR filter with 31 taps yields a filter characteristic that is

Figure 11. Functional logic diagram of bit-serial, finite-impulse-response approximation of a third-order Butterworth digital filter.

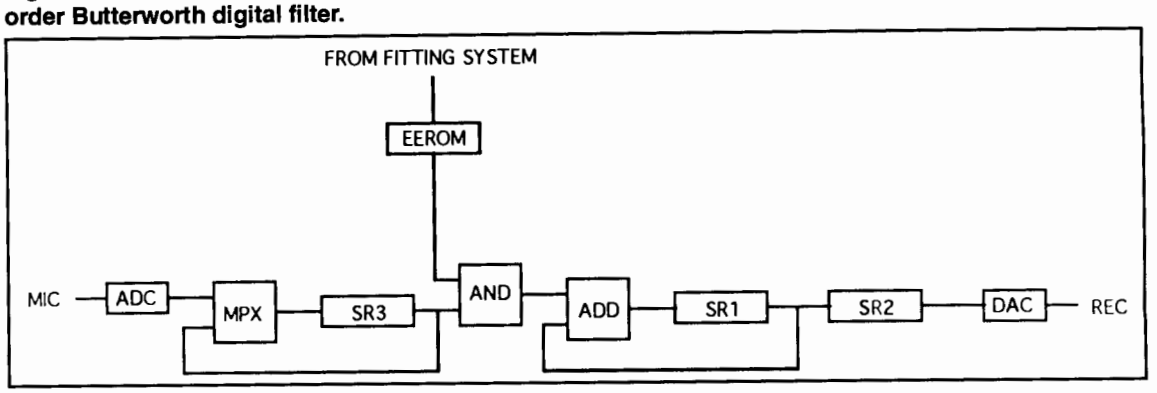


Table 5. Summary of finite-impulse-response filter design (number of devices).

Device	Number of stages	Transistors per stage	Total number of transistors	Area $\lambda = 1\ \mu m$
SR1	30	14	420	
SR2	14	14	196	
SR3	31x14=434	14	6,076	
ADDER	1	40	40	
Total			6,732	6.0588 mm^2

Table 6. Summary of finite-impulse-response filter design (state transitions and power).

Device	Shifts per sample period	Transistor state changes per sample period	Switching power $\lambda = 1\ \mu m$ fs=20 kHz
SR1	30x14x31=13,020	13,020x420=5,468,400	
SR2	14	14x196=2,744	
SR3	31x14=434	434x6,076=2,636,984	
ADDER	13,020	13,020x40=520,000	
Totals		8,628,928	868 µW

within 0.3 dB from 0 to 1 kHz, falls to 30 dB at 2 kHz, and has an out-of-band rejection of 30 dB for frequencies above 2 kHz. Therefore, a 31-tap FIR filter should do nicely.

The expression for the FIR filter is:

$$y(n) = \sum c_i * x(n-i)$$

where ci are the filter coefficients and x(n-i) are input values delayed by index, i. Note that if the filter is symmetric, its phase response will be linear. In that case, the coefficients will be symmetric about the center tap, and the delay through the filter will be 15 samples or 750 µs assuming a sampling rate of 20 kHz. If we assume that the coefficients are 10 bits, that the samples are 14 bits, and that the number of guard bits to prevent accumulator overflow is 6, the design criteria for the structure in Figure 11 is as follows: Shift register, SR3 is 434 bits long, SR1 is 30 bits long, and SR2 is 14 bits long. To compute each output sample of the filter requires that SR3 is shifted 434 times, SR2 is shifted 14 times, and SR1 is shifted 930 times per sample period. The number of transistors and area of this circuit are summarized in Table 5. The total number of bit transitions for each element and the number of transistors involved per sample are summarized in Table 6. Using the approximations described previously, the total switching power will be about 868 µw with a 1 µm feature size and the area will be about 6.1 mm^2.

We have seen that it is possible to implement a digital filter that can simulate the characteristic of an analog switched-capacitor filter. In either approach (analog or digital) the power dissipation and size of the circuits are compatible with hearing aid batteries and ear level packaging. However, the advantages of digital implementation become greater as we attempt to achieve greater miniaturization or greater functionality.

Discussion

In reviewing the current state-of-the-art of hearing aid design, it is clear that hearing aid technology has advanced significantly over the past decade. The weakest component would appear to be the signal processing circuits, which are currently limited to relatively simple filters and compression circuits. Because of the rapid advancements that have been made in the technology of fabricating smaller and smaller VLSI circuitry, it is likely that the major advances in hearing aids will be in the area of signal processing complexity. It has been argued above (and demonstrated) that digital processing can provide this additional complexity and that this new technology is now practical to implement in an ear level hearing aid form. It has also been shown above that relatively modest digital processing structures can provide a versatility that is difficult to achieve with analog designs.

With regard to functionality, our crystal ball is cloudy about what features provide benefit to the listener with impaired hearing. However, certain design criteria can be stated. The hearing aid should have a wide dynamic range. It should not saturate for loud signals, which are often those conditions that are the noisiest. If a hearing aid saturates for a signal corrupted by noise, the signal and noise spectra are irretrievably distorted in a way that even a normal ear cannot sort out. Therefore, one emphasis should be on high fidelity even though the hearing impaired listener may not be able to directly appreciate the signal quality.

Second, everyone seems to agree that audibility is important to speech intelligibility. This would argue that being able to shape the spectrum of the signal to compensate for loss of sensitivity is important. It is not clear what degree of precision is required for spectral shaping. However, it is unlikely that conventional hearing aids that utilize only a small number of channels or that utilize filters with only a few programmable cutoff frequencies will provide sufficient adjustment flexibility to fit a wide enough variety of individuals. On the other hand, a relatively short digital FIR filter with programmable coefficients can provide the greater adjustment flexibility.

Another aspect of audibility is masking of signals at supra-threshold levels. A number of investigators have observed that both upward and downward spread of masking occurs in the cochlea, that these masking functions may be different for the impaired ear, and that there are considerable individual differences in the masking characteristics. Therefore, one might argue that an optimum fit should take masking relations into account and that when this is done, the need for careful shaping of the signal spectrum on an individualized basis may be more important than it is now believed to be. In addition, an optimal fit may be a function of level, which brings us to the topic of compression.

Traditional compression circuits provide a means for reducing the range of input signals to better match the residual dynamic range of the impaired ear. Multi-channel compression hearing aids are an attempt to provide different compression functions for different ranges of frequencies. It is clear that the design trend is to make compression functions more complicated and provide a greater degree of flexibility with regard to compression ratios and thresholds. Another approach to compression that is relatively new (Killion, 1990) utilizes two transfer functions, one that is active at low levels and another that is active at high signal levels. This provides for gain and spectral shaping at low levels and a transparent hearing aid characteristic at high levels. The design trend here is obvious as we see greater effort focused on developing more sophisticated compression functions. As a result of this trend, one can envision future hearing aids that transform signals into the useable range of the impaired ear in complicated and precise ways and do it automatically. Digital processing is a likely candidate for implementing these more sophisticated compression algorithms.

Finally, there are a number of advantages with regard to digital hearing aid circuits that cannot be measured in terms of direct benefit for the hearing impaired user. This has to do with manufacturability, reliability, and testability. It can be expected that as more and more analog circuitry is replaced by digital circuitry that chip yields will increase, costs will decrease, and that frequency of repairs will be lower.

Address all correspondence to: A. Maynard Engebretson, D.Sc., Applied Research Laboratory, Department of Computer Science, Washington University, One Brookings Drive, St. Louis, MO 63130 USA

References

Callius, F., Salchli, F.H., & Girard, D. (1989). A set of 4 IC's in CMOS technology for a programmable hearing aid. *IEEE Journal of Solid-State Circuits, Vol. 24*, (2), 301-312.

Candy, J.C., & Temes, G.C. (1992). *Oversampling delta-sigma data converters*. New York, NY: IEEE Press.

Daniels, R.W. (1974). *Approximation methods for electronic filter design*. McGraw-Hill.

Dyrlund, O., & Bisgaard, N. (1991). Acoustic feedback margin improvements in hearing instruments using a prototype DFS (Digital Feedback Suppression) system. *Scand Audiol, 20*, 49-53.

Egolf, D.P., & Larson, V.D. (1984). Acoustic Feed back Suppression in Hearing Aids (163-164). (Department of Veteran Affairs Rehabilitation R&D Progress Reports, 1984, Publication Section (110A1).) Washington, DC: U.S. Government Printing Office.

Egolf, D.P., & Larson, V.D. (1986). Studies of Acoustic Feedback in Hearing Aids (315). (Department of Veteran Affairs Rehabilitation R&D Progress Reports, 1986, Publication Section (110A1).) Washington, DC: U.S. Government Printing Office.

Engebretson, A.M., Morley, R.E., & Popelka, G.R. (1987). Development of an ear-level digital hearing aid and computer-assisted fitting procedure. *J. Rehab. Res. Devel., 24*(4), 55-64.

Engebretson, A.M., O'Connell, M.P., & Gong, F. (1990). An adaptive feedback equalization algorithm for the CID digital hearing aid (2286-2287). New York: IEEE Publishing Services.

Engebretson, A.M., French-St. George, M., & O'Connell, M.P. (1991, June). Adaptive feedback stabilization of hearing aids. The Second International Workshop on Hearing Impairment and Signal-Processing Hearing Aids, London, England, June 1991.

Engel, G.L. (1990). Design and Analysis of logarithmic digital quantizers with applications to low-power data interfaces for speech processing. Unpublished doctoral dissertation, Department of Electrical Engineering, Sever Institute of Technology, Washington University, St. Louis, MO, May 1990.

French-St. George, M., & Metzger, H. (1990). Fitting the CID digital hearing aid: Accuracy and subjective frequency/gain preferences. *Issues in Advanced Hearing Aid Research*, House Ear Institute, Lake Arrowhead, CA, June 1990.

French-St. George, M., Engebretson, A.M., & O'Connell, M.P. (1992). Behavioral assessment of CID's benchtop (version 2) digital hearing aid (DHA): Noise Reduction. *J. Amer. Academy Audiol., 3*(2), 132-141, March 1992.

Harris, F J. (1978) On the use of windows for harmonic analysis with the discrete Fourier transform. *Proceeding of the IEEE, 66*(1), 51-83, January 1978.

Haskard, M.R., & May, I.C. (1988). *Analog VLSI design.* New York: Prentice Hall.

Kaiser, J.F. (1974). Nonrecursive filter design using the I0-SINH window function. Proceedings of the 1974 IEEE Int. Symp. on Circuits and Syst. (20-23). New York: IEEE Press.

Killion, M.C. (1976). Noise of ears and microphones. *J. Acoust. Soc. Am., 59* (2), 424-433.

Killion, M.C. (1990). A high fidelity hearing aid. *Hearing Instruments, 41*(8), 38-39.

Lim, J.S., & Oppenheim, A.V. (1979). Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE, 67*(12), 1586-1604.

Mead, C., & Conway, L. (1980). *Introduction to VLSI systems.* Reading, MA.: Addison-Wesley.

National Research Council. (1989). *Removal of noise from noise-degraded speech signals.* Committee on Hearing, Bioacoustics, and Biomechanics. National Academy Press: Washington, DC.

Rabiner, L.R., & Schafer, R.W. (1978). *Digital processing of speech signals.* Englewood Cliffs, NJ.: Prentice-Hall.

Rankovic, C.M., Freyman, R.L., & Zurek, P.M. (1992). Potential benefits of adaptive frequency-gain characteristics for speech reception in noise. *J. Acoust. Soc. Am., 91*(1), 354-362.

Widrow, B., Glover, J.R., McCool, J.M., Kaunitz, J., Williams C.S., Hearn, R.H., Ziedler, J.R., Dong, E., & Goodlin, R.C. (1975). Adaptive noise cancelling: Principles and applications. *Proceedings of the IEEE, 63*(12), 1692-1716.