LIMITS FOR GENERALIZED SIDELOBE CANCELLERS WITH EMBEDDED ACOUSTIC ECHO CANCELLATION

W. Herbordt and W. Kellermann

University of Erlangen-Nuremberg, Telecommunications Laboratory
Cauerstr. 7, D-91058 Erlangen, Germany
{ herbordt, wk}@LNT.de

ABSTRACT

This paper analyzes positive synergies and theoretical limits of the combination of acoustic echo cancellation (AEC) and Generalized Sidelobe Cancellers (GSC) for the removal of echoes of loudspeaker signals and local interferers. While the proposed system only requires one AEC for an arbitrary number of microphones, the array gain is limited by the number of sensor channels when all interferers arrive from different directions-of-arrival (DOAs). The paper also shows that the degrees of freedom of the adaptive sidelobe cancelling path are not sufficient when local interferers and acoustic echoes have common DOAs.

1. INTRODUCTION

Natural communication between the user and personal computing devices calls for speech-driven application control via hands-free acoustical interfaces. The user may thus move freely without wearing or holding any microphone device. For optimum speech communication quality and at maximum speech recognition rates, such human/PC interfaces need to produce signals of interest that a) are free of background noise, reverberation, and echoes of loudspeaker signals, and that b) can be distinguished from interfering local speakers.

In contrast to desktop PCs for mobile personal devices, the computational power of mobile personal devices will remain a limiting factor in the usage of complex noise-reduction algorithms. Compared to single-channel temporal filtering noise-reduction schemes, multichannel space-time filtering promises better target signal quality and more efficient suppression of interferers (local interferers).

AEC is desirable whenever a reference of the interference is accessible. With personalized devices, these interferers may be echoes from the loudspeakers that are part of the device (acoustic echoes).

This research aims at reconciling multichannel noise-reduction techniques and AEC while preserving the efficiency of both approaches and keeping the computational load moderate. In [1], three basic concepts have been presented. These methods have been applied to the combination of a robust GSC [2] and AEC in [3]. It has been shown that placing one AEC in the reference path behind the fixed beamformer (FBF) seems to be most promising. Only one AEC is required for an arbitrary number of microphones (see Fig. 1). However, when acoustic echoes are efficiently suppressed by the AEC, they may still leak through the sidelobe cancelling path that is formed by the adaptive blocking matrix (ABM) and the adaptive interference canceller (AIC) due to the limited number of degrees of freedom of the adaptive filters (AF) in the AIC.

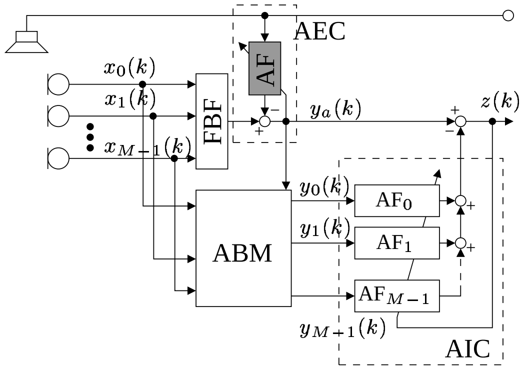

Fig. 1. AEC embedded into GSC: GSAEC

This paper shows theoretical limits of the interference rejection of GSAEC in anechoic environments with a finite number of interferers as well as in diffuse noise fields. First, the optimization problem of finding AIC filters that maximize the array gain is formulated (Section 2). In Section 3, the solution of this equation is discussed for variable number of interferers with the assumption of single-path propagation. In Section 4, the optimization problem is applied to diffuse noise fields, and theoretical limits of local interference rejection and acoustic echo rejection are derived.

2. PROBLEM FORMULATION

Throughout the paper, the array gain [4] is extended to the combination of beamformer and AEC: The SNR improvement with respect to acoustic echoes and with respect to

This work was supported by a grant from Intel Corp., Hillsboro, OR.

local interferers are termed echo rejection and local interference rejection, respectively.

Our analysis is based on the discrete-time Fourier transform. In an arrangement with L discrete interferers and M microphones, interfering signals are described by $X_l(\omega)$ with $l = 0, 1, \ldots, L-1$. $X_{l,i}(\omega)$ describes the contribution of the l -th interferer to the i-th sensor channel, with $i = 0, 1, \ldots, M-1$. For now, we do not distinguish local interferers from acoustic echoes.

All noise contributions should be stationary and ergodic ao that cross power spectral densities may thus be written as

$$\Phi_{l,ij}(\omega) = X_{l,i}(\omega)^* X_{l,j}(\omega)\,, \tag{1}$$

where * denotes complex conjugation. Furthermore, they should be spatially homogeneous, i.e.

$$\Phi_l(\omega) = \Phi_{l,ii}(\omega) = \Phi_{l,jj}(\omega) \quad \forall\, i, j\,, \tag{2}$$

and they should be mutually orthogonal, or

$$X_{k,i}(\omega)^* X_{l,j}(\omega) = 0 \quad \forall\, i,\, j;\, k \neq l\,. \tag{3}$$

Using the normalized complex spatial coherence function $\rho_{l,ij}(\omega)$

$$\rho_{l,ij}(\omega) = \frac{\Phi_{l,ij}(\omega)}{\sqrt{\Phi_{l,ii}(\omega)\Phi_{l,jj}(\omega)}}\,, \tag{4}$$

Eq. 1 can be written as

$$\Phi_{l,ij}(\omega) = \Phi_l(\omega)\rho_{l,ij}(\omega)\,. \tag{5}$$

The response vector of FBF is given by the frequency responses of each channel filter $W_i(\omega)$

$$\mathbf{w}^T(\omega) = \left(W_0(\omega)\ W_1(\omega)\ \cdots W_{M-1}(\omega)\right). \tag{6}$$

The AEC only attenuates acoustic echoes without distorting local interferers. The suppression of the AEC of the l-th interference is described by the inverse gain $1/G_{a,l}(\omega)$, $|G_{a,l}(\omega)| \in [0; 1]$. Here, local interferers have to be treated differently than acoustic echoes: $G_{a,l}(\omega) = 1$ for local interferers, and $|G_{a,l}(\omega)| \in [0; 1]$ for acoustic echoes. The AEC output $Y_a(\omega)$ can be written as

$$Y_a(\omega) = \sum_{l=0}^{L-1}\sum_{i=0}^{M-1} W_i(\omega) G_{a,l}(\omega) X_{l,i}(\omega)\,. \tag{7}$$

The ABM is assumed to be ideal, that is, all interferers are passed without being distorted while the desired signal is perfectly suppressed. The i-th ABM output is then given by:

$$Y_i(\omega) = \sum_{l=0}^{L-1} X_{l,i}(\omega)\,. \tag{8}$$

Capturing the ABM output signals $Y_i(\omega)$ and the AIC filter transfer functions $H_i(\omega)$ in vectors

$$\mathbf{y}^T(\omega) = \left(Y_0(\omega)\ Y_1(\omega)\ \cdots\ Y_{M-1}(\omega)\right), \tag{9}$$

and

$$\mathbf{h}^T(\omega) = \left(H_0(\omega)\ H_1(\omega)\ \cdots\ H_{M-1}(\omega)\right), \tag{10}$$

respectively, the GSAEC output signal $Z(\omega)$ can be written as

$$Z(\omega) = Y_a(\omega) - \mathbf{h}^T(\omega)\mathbf{y}(\omega)\,. \tag{11}$$

Thus, to optimize the array gain, the AIC filters have to solve the least-squares problem

$$|\mathbf{h}^T(\omega)\mathbf{y}(\omega) - Y_a(\omega)|^2 \rightarrow \min_{\mathbf{h}(\omega)}\,. \tag{12}$$

3. POINT SOURCES WITH SINGLE-PATH PROPAGATION

Here, the solution of the least-squares problem Eq. 12 is discussed for an arbitrary number of interferers with a single propagation path each. The propagation delay between the l-th interference and the i-th microphone is denoted by τ_{li}. The contribution of the l-th interferer to the i-th sensor signal is thus given by

$$X_{li}(\omega) = X_l(\omega)\exp(-j\omega\tau_{li})\,. \tag{13}$$

Introducing matrices

$$\mathbf{Y}(\omega) = \{X_l(\omega)\exp(-j\omega\tau_{li})\}_{L\times M} \tag{14}$$

and

$$\mathbf{Y}_g(\omega) = \{G_{a,l}X_l(\omega)\exp(-j\omega\tau_{li})\}_{L\times M}\,, \tag{15}$$

and using the assumption of mutually orthogonal interferers, the minimization problem Eq. 12 can be written as follows:

$$Z_{min}(\omega) = \min_{\mathbf{h}(\omega)} ||\mathbf{Y}_g(\omega)\mathbf{w}(\omega) - \mathbf{Y}(\omega)\mathbf{h}(\omega)||^2\,. \tag{16}$$

Two trivial cases with $Z_{min}(\omega) = 0$ can be considered: First, without echo cancellation, i.e. $G_a(\omega) = 1$, ideal interference rejection is given for $\mathbf{h}_{opt}(\omega) = \mathbf{w}(\omega)$. Second, with perfect echo cancellation, i.e. $G_a(\omega) = 0$, and without local sources, all interferers are ideally suppressed for $\mathbf{h}_{opt}(\omega) = 0$. Depending on the rang of the matrix $\mathbf{Y}(\omega)$, these trivial cases can be found in different classes that are discussed in the following.

In the general case, we observe that ideal interference rejection $Z_{min}(\omega) = 0$ is obtained, if[1]

$$\mathbf{Y}(\omega)\mathbf{h}(\omega) = \mathbf{Y}_g(\omega)\mathbf{w}(\omega) \quad \forall\,\omega\,. \tag{17}$$

This system of linear equations can be solved, if

$$\text{rank}\left\{\left(\mathbf{Y}(\omega), \mathbf{Y}_g(\omega)\mathbf{w}(\omega)\right)_{L\times M+1}\right\} = \text{rank}\left\{\mathbf{Y}(\omega)_{L\times M}\right\}. \tag{18}$$

If $\text{rank}\{\mathbf{Y}(\omega), \mathbf{Y}_g(\omega)\mathbf{w}(\omega)\} \neq \text{rank}\{\mathbf{Y}(\omega)\}$, Eq. 17 cannot be solved, and the optimal value of $\mathbf{h}(\omega)$ is found by setting the partial derivative of $||Z(\omega)||^2$ with respect to $\mathbf{h}(\omega)$ equal to zero. That is,

$$\mathbf{Y}^H(\omega)\mathbf{Y}(\omega)\mathbf{h}(\omega) = \mathbf{Y}^H(\omega)\mathbf{Y}_g(\omega)\mathbf{w}(\omega)\,, \tag{19}$$

where ^{H} denotes the Hermitian operator.

[1] The matrix $\left(\mathbf{Y}(\omega), \mathbf{Y}_g(\omega)\mathbf{w}(\omega)\right)$ is obtained by appending the column vector $\mathbf{Y}_g(\omega)\mathbf{w}(\omega)$ to the right side of the matrix $\mathbf{Y}(\omega)$.

Depending on the rank R and the dimensions of the $L \times M$ matrix $\mathbf{Y}(\omega)$, four classes of solutions for the minimization problem must be distinguished:

(a) $R = L$: (a1) $L = M$: There are as many interferers as microphones with a different DOA each. The system of linear equations Eq. 17 is solved by

$$\mathbf{h}(\omega) = \mathbf{Y}^{-1}(\omega)\mathbf{Y}_g(\omega)\mathbf{w}(\omega)\,,$$

and consequently $Z_{min}(\omega) = 0$.

(a2) $L < M$: There are less interferers than microphones with a different DOA each. The matrix $\mathbf{Y}(\omega)$ is of full row rank and it can be shown that one solution of the underdetermined least-squares problem is given by [5]

$$\mathbf{h}_{opt}(\omega) = \mathbf{Y}^H(\omega)\left(\mathbf{Y}(\omega)\mathbf{Y}^H(\omega)\right)^{-1}\mathbf{Y}_g(\omega)\mathbf{w}(\omega)\,. \quad (20)$$

It may be easily verified that Eq. 17 is fulfilled by substituting Eq. 20 into $||Z(\omega)||^2$. Optimum interference rejection is thus possible.

(b) $R = M$ and $L > M$: The number of interferers with different DOAs is greater than the number of microphones. $\mathbf{Y}(\omega)$ has full column rank, and the matrix $(\mathbf{Y}^H(\omega)\mathbf{Y}(\omega))$ is invertible in turn. With Eq. 19, the optimum AIC filters $\mathbf{h}(\omega)$ can thus be determined as follows:

$$\mathbf{h}_{opt}(\omega) = \left(\mathbf{Y}^H(\omega)\mathbf{Y}(\omega)\right)^{-1}\mathbf{Y}^H(\omega)\mathbf{Y}_g(\omega)\mathbf{w}(\omega)\,. \quad (21)$$

Substituting Eq. 21 into $||Z(\omega)||^2$ gives the optimum interference rejection $Z_{min}(\omega)$. That is,

$$Z_{min}(\omega) = \left(\mathbf{Y}_g(\omega)\mathbf{w}(\omega)\right)^H\left(\mathbf{Y}_g(\omega)\mathbf{w}(\omega) - \mathbf{Y}(\omega)\mathbf{h}(\omega)\right)\,. \quad (22)$$

$Z_{min}(\omega) = 0$ is only given, if exactly $L - R$ interferers with common $G_a(\omega)$ arrive from same DOAs.

(c) $R < M < L$ or $R < L \leq M$: Here, no closed form of optimum filter coefficients can be given. However, a solution in the least-squares sense is found by applying singular value decomposition [5]. That is,

$$\mathbf{h}_{opt}(\omega) = \mathbf{U}(\omega)\begin{pmatrix} \mathbf{S}^{-1}(\omega) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{V}^H(\omega)\mathbf{Y}_g(\omega)\mathbf{w}(\omega)\,.$$

The unitary matrices $\mathbf{U}(\omega)$ and $\mathbf{V}(\omega)$ are determined such that

$$\mathbf{V}^H(\omega)\mathbf{Y}(\omega)\mathbf{U}(\omega) = \begin{pmatrix} \mathbf{S}(\omega) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\,,$$

with the matrix $\mathbf{S} = \mathrm{diag}\left(\sigma_0(\omega)\ \sigma_1(\omega)\cdots\sigma_{R-1}(\omega)\right)$.

$Z_{min}(\omega) = 0$, i.e. Eq. 18 is fulfilled, if exactly $L - R$ interferers with common $G_a(\omega)$ arrive from same DOAs.

In summary, ideal suppression of acoustic echoes and local interferers is possible when the acoustic echoes arrive from different DOAs than the local interferers and when the total number of interferers from different DOAs is less or equal to the number of microphones.

This means that in real applications positive synergies between AEC and GSC may be expected in environments with few interferers and with acoustic echoes arriving from different DOAs than local interferers. Otherwise, better interference rejection may be obtained by applying GSC only.

4. DIFFUSE NOISE FIELDS

In the following section, upper and lower limits are derived for diffuse interfering wave fields. This corresponds to the case of an infinite number of correlated interferers. Without loss of generality, presence of a single local interferer and presence of a single echo signal is assumed for simplicity. First, the optimum AIC filters in the Minimum Mean-Squared Error (MSE) sense are derived. Then, expressions for the acoustic echo rejection and local interference rejection are developed. Finally, upper and lower limits are discussed.

The optimum AIC filters $\mathbf{h}_{opt}(\omega)$ are given by [5]:

$$\mathbf{h}^T_{opt}(\omega) = \mathbf{P}^{-1}(\omega)\mathbf{r}_{Y_aY}(\omega)\,, \quad (23)$$

with the cross spectral density matrix

$$\mathbf{P}(\omega) = \left\{\Phi_{Y_iY_j}(\omega)\right\}_{M\times M}\,, \quad (24)$$

where i is the column index and j is the row index, and with the cross spectral density vector

$$\mathbf{r}_{Y_aY}(\omega) = \left\{\Phi_{Y_aY_i}(\omega)\right\}_{M\times 1}. \quad (25)$$

The power spectral densities (PSD) of the local interferer and of the acoustic echo signal are given by $\Phi_{X_IX_I}(\omega)$ and $\Phi_{X_EX_E}(\omega)$, respectively. The spatial coherence of the assumed diffuse noise fields $\rho_{ij}(\omega)$ is given by [4]

$$\rho_{ij}(\omega) = \mathrm{sinc}\left(\frac{\omega d(i-j)}{c}\right)\,, \quad (26)$$

when a linear array with sensor spacing d is assumed. c denotes the propagation speed. In the following, the spatial coherence functions between the sensors after the FBF filters $\mathbf{w}(\omega)$ will be combined in a matrix

$$\mathbf{R}(\omega) = \{\rho_{ij}(\omega)\}_{M\times M}\,, \quad (27)$$

where i is the column index and j is the row index. Assuming mutually uncorrelated local interference and acoustic echo, Eq. 24 and Eq. 25 may then be computed as

$$\mathbf{P}(\omega) = \left(\Phi_{X_IX_I}(\omega) + \Phi_{X_EX_E}(\omega)\right)\mathbf{R}(\omega)\,, \quad (28)$$

and

$$\mathbf{r}^T_{Y_aY}(\omega) = \left(\Phi_{X_IX_I}(\omega) + G_a(\omega)\Phi_{X_EX_E}(\omega)\right)\cdot \underbrace{\sum_{j=0}^{M-1} W_j(\omega)\left(\rho_{j0}(\omega)\ \rho_{j1}(\omega)\ \cdots\ \rho_{j,M-1}(\omega)\right)}_{:=\mathbf{r}^T(\omega)}\,, \quad (29)$$

respectively. In the absence of AEC, Eq. 23 reduces to $\mathbf{h}(\omega) = \mathbf{R}^{-1}(\omega)\mathbf{r}(\omega)$ when Eqs. 28, 29 are observed. With Eqs. 7, 11, the optimum coefficients are given by $\mathbf{h}_{opt}(\omega) = \mathbf{w}(\omega)$. Finally, the general case reads with Eqs. 23, 28, 29:

$$\mathbf{h}_{opt}(\omega) = \frac{1 + G_a(\omega)\eta(\omega)}{1 + \eta(\omega)}\mathbf{w}(\omega)\,, \quad (30)$$

with $\eta(\omega) = \Phi_{X_E X_E}(\omega)/\Phi_{X_I X_I}(\omega)$ the ratio of power spectral densities. With Eq. 7 through Eq. 11, $\Phi_{ZZ}(\omega) = Z(\omega)Z^*(\omega)$ may be written as follows:

$$\Phi_{ZZ}(\omega) = \left(\eta^2(\omega)\Phi_{X_I X_I}(\omega) + \Phi_{X_E X_E}(\omega)\right) \cdot \cdot \frac{|1 - G_a(\omega)|^2}{\left(1 + \eta(\omega)\right)^2} \rho(\omega) , \tag{31}$$

where

$$\rho(\omega) = \sum_{i=0}^{M-1} \sum_{j=0}^{M-1} W_i(\omega) W_j^*(\omega) \rho_{ij}(\omega) . \tag{32}$$

We see that $\left(\frac{1}{2\pi}\int_0^{2\pi} \rho(\omega) d\omega\right)^{-1}$ is equivalent to the array gain of FBF when the response vector $\mathbf{w}(\omega)$ is normalized such that the array response is unity for the DOA of the target.

We now split $\Phi_{ZZ}(\omega)$ in Eq. 31 into terms with local interferer components and into terms with acoustic echo components $\Phi_{Z_x Z_x}(\omega)$, $x \in \{I, E\}$:

$$\Phi_{ZZ}(\omega) = \Phi_{Z_I Z_I}(\omega) + \Phi_{Z_E Z_E}(\omega) . \tag{33}$$

Then, we decompose the local interference and the acoustic echoes into narrow frequency intervals of width $2\Delta\omega$ around the center frequency ω_0, and we define

$$G_x(\omega_0) = \frac{\int_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \Phi_{Z_x Z_x}(\omega)\, d\omega}{\int_{\omega_0-\Delta\omega}^{\omega_0+\Delta\omega} \Phi_{X_x X_x}(\omega)\, d\omega} . \tag{34}$$

The array gain with respect to local interferers or with respect to acoustic echoes is thus given by $1/G_x(\omega_0)$.

$G_I(\omega_0)$ and $G_E(\omega_0)$ can be written as:

$$G_I(\omega_0) = \frac{\eta(\omega_0)|1 - G_a(\omega_0)|^2}{\left(1 + \eta(\omega_0)\right)^2} \rho(\omega_0) , \tag{35}$$

and

$$G_E(\omega_0) = \frac{|1 - G_a(\omega_0)|^2}{\left(1 + \eta(\omega_0)\right)^2} \rho(\omega_0) , \tag{36}$$

respectively. Several special cases can be considered. Generally, ideal suppression is possible for $G_a(\omega_0) = 1$. Then, assume that either no acoustic echo signal or no local interferer is present. It follows that $\eta(\omega_0) = 0$ or $\eta(\omega_0) = +\infty$ and ideal suppression $G_I(\omega_0)$ and $G_E(\omega_0)$ is possible, respectively.

The worst case is obtained when both interferers have identical energy, or $\eta(\omega_0) = 1$, and the AEC ideally suppresses acoustic echoes, or $G_a(\omega_0) = 0$. Then,

$$G_E(\omega_0) = G_I(\omega_0) = \frac{\rho(\omega_0)}{4} . \tag{37}$$

This corresponds to 6 dB array gain relative to FBF.

In Fig. 2(a)(b), the interference rejection $1/G_I(\omega_0)$ and the acoustic echo rejection $1/G_E(\omega_0)$ are illustrated as a function of ω_0 with parameter $\eta(\omega_0) \in \{0.01, 1, 10\}$, respectively. The array consists of $M = 8$ equally spaced sensors that are arranged linearly with a distance $d = 4$ cm. FBF is a simple delay&sum beamformer, or $w_i(\omega) = 1/M$.

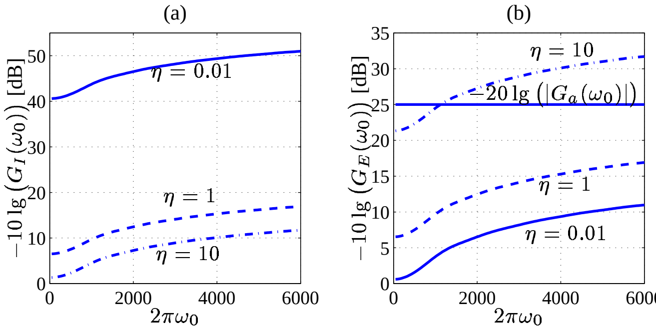

Fig. 2. Local interference rejection(a) and acoustic echo rejection(b) for $-20\lg\left(|G_a(\omega_0)|\right) = 25$ dB over frequency

It can be seen that when the acoustic echoes are 10 dB stronger than the local interferer $\left(\eta(\omega_0) = 10\right)$, then $-10\lg\left(G_I(\omega_0)\right)$ is reduced to less than 12 dB. Moreover, an acoustic echo rejection greater than that of the AEC $\left(G_E(\omega_0) > |G_a(\omega_0)|\right)$ can only be obtained for $2\pi\omega_0 > 1200$ Hz. On the other hand, when $\eta(\omega_0) = 0.01$, more than 40 dB local interference rejection is reached, but the acoustic echo leakage through the sidelobe cancelling path is between 14 dB and 24 dB.

5. CONCLUSION

In this work, we have investigated synergies of AEC embedded into GSC. We have shown that optimum interference rejection is only possible when local interferers and acoustic echoes have no DOAs in common and when the total number of interferers is less than the number of microphones. The lowest array gain is obtained for diffuse noise fields with identical PSD of local interferers and acoustic echoes: The array gain is only 6 dB higher than that of FBF.

6. REFERENCES

[1] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays," *IEEE Proc. ICASSP*, vol. 1, pp. 219–222, April 1997.

[2] O. Hoshuyama, A. Sugiyama, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. on SP*, vol. 47, no. 10, October 1999.

[3] W. Herbordt, W. Kellermann, "Acoustic echo cancellation embedded into the generalized sidelobe canceller," *EURASIP Proc. EUSIPCO*, vol. 3, September 2000.

[4] D.H. Johnson, D.E. Dudgeon, *Array Signal Processing*, Prentice Hall, Englewood Cliffs, 1993.

[5] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, Upper Saddle River, New Jersey, 3rd edition, 1996.