MISSOURI S&T

Missouri University of Science and Technology

Scholars' Mine

Electrical and Computer Engineering Faculty Research & Creative Works

Electrical and Computer Engineering

01 Jan 1999

Double-Talk Robust Fast Converging Algorithms for Network Echo Cancellation

T. Gansler

Steven L. Grant
Missouri University of Science and Technology, sgrant@mst.edu

J. Benesty

M. M. Sondhi

Follow this and additional works at: https://scholarsmine.mst.edu/ele_comeng_facwork

Part of the Electrical and Computer Engineering Commons

Recommended Citation

T. Gansler et al., "Double-Talk Robust Fast Converging Algorithms for Network Echo Cancellation," *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 1999*, Institute of Electrical and Electronics Engineers (IEEE), Jan 1999.
The definitive version is available at https://doi.org/10.1109/ASPAA.1999.810888

DOUBLE-TALK ROBUST FAST CONVERGING ALGORITHMS FOR NETWORK ECHO CANCELLATION

Tomas Gänsler

Steven L. Gay
M. Mohan Sondhi
Jacob Benesty

Dept. of Applied Electronics
Lund University
P.O. Box 118, 221 00 Lund, Sweden
tg@tde.lth.se

Bell Laboratories, Lucent Technologies
600 Mountain Avenue
Murray Hill, New Jersey 07974-0636
{slg,mms,jbenesty}@bell-labs.com

ABSTRACT

Echo cancelers which cover longer impulse responses ($\geq$ 64 ms) are desirable. Long responses create a need for more rapidly converging algorithms in order to meet the specifications for network echo cancelers devised by ITU (International Telecommunication Union). In general, faster convergence implies a higher sensitivity to near-end disturbances, especially "double-talk." Recently, a fast converging algorithm called *Proportionate NLMS* (Normalized Least Mean Squares) algorithm (PNLMS) has been proposed. This algorithm exploits the sparseness of the echo path. In this paper we propose a method for making the PNLMS algorithm more robust against double-talk. The slower divergence rate of these robust algorithms in combination with a standard Geigel double-talk detector improves the performance of a network echo canceler considerably during double-talk. This results in the robust PNLMS algorithm which diverges much slower than PNLMS and standard NLMS. A generalization of the robust PNLMS algorithm to a robust proportionate Affine Projection Algorithm (APA) is also presented. It converges very fast, and unlike PNLMS, is not as dependent on the assumption of a sparse echo path response. Trade off between convergence and divergence rate is easily tuned with one parameter and the added complexity is about 7 instructions per sample.

1. INTRODUCTION

There is a need for network echo cancelers for echo paths with long impulse responses ($\geq$ 64 ms). However, longer impulse responses slow down the convergence rate, [1, 2], thus rendering traditional algorithms like NLMS inadequate. It will therefore, be desirable to implement fast-converging algorithms in future echo cancelers. In [2, 3], faster converging algorithms called Proportionate NLMS (PNLMS) and PNLMS++ respectively are proposed. These algorithms achieve higher convergence rate by using the fact that the active part of a network echo path is usually much smaller (4–8 ms) compared to 64–128 ms of the whole echo path that has to be covered.

Besides convergence rate and complexity issues, an important aspect of an echo canceler is its performance during "double-talk" (i.e. simultaneous far- and near-end speech). A high convergence rate is usually accompanied by a high divergence rate in the pres-

ence of double-talk. This mode in a conversation perturb the adaptive filter of the echo canceler (EC) so that it does not attenuate the echo sufficiently. To inhibit the divergence of the EC the standard procedure is to use a level based double-talk detector (DTD) [4]. Whenever double-talk is detected the step-size of the adaptive filtering algorithm is set to zero thus inhibiting the adaptation. *Unfortunately, during the time required by the DTD to detect double-talk, the echo canceler often diverges. This is because a few (e.g. < 5) undetected large amplitude samples perturb the echo path estimate considerably.*

This work focuses on how to decelerate the divergence of algorithm due to undetected double-talk while maintaining good convergence rate of the canceler. Our approach has its roots in the theory of robust statistics and is based on introducing a scaled nonlinearity into the adaptive algorithm. The nonlinearity limits the impact of large disturbances on the coefficient setting. This idea was developed for a subband echo canceler in [5] and showed promising results. However, neither the trade off between robustness and convergence rate nor the performance for fullband adaptive filters were studied in that paper. The robust algorithm developed here combines the PNLMS++ algorithm with the appropriate nonlinearity.

The robust PNLMS++ algorithm is also generalized to a robust Proportionate Affine Projection Algorithm (PAPA), – an algorithm which is a combination of Affine Projection Algorithm, [6], and the proportionate step-size technique.

2. ADAPTIVE ALGORITHMS

In derivations and descriptions the following notations are used, see also Fig. 1.

The excitation vector is denoted $\mathbf{x}_n = [x_n, \ldots, x_{n-L+1}]^T$ where x_n is the far-end speech signal. v_n is the background noise and w_n is the near-end speech (double-talk). The near-end signal, i.e. echo and noise possibly near-end speech, is denoted y_n. The residual echo is $e_n = y_n - \mathbf{h}_n^T \mathbf{x}_n$ where $\mathbf{h}_n = [h_{0,n}, \ldots, h_{L-1,n}]^T$ is the estimated echo path. Here L is the length of the adaptive filter. The PNLMS algorithm was proposed in [2]. For line echo cancellation, it is reasonable to assume that the echo path is *sparse*, i.e. many coefficients are zero, and try to identify only the non-zero active coefficients. This is the idea behind the PNLMS algorithm which is a modification of the NLMS algorithm. In PNLMS, an adaptive individual step-size is assigned to each filter coefficient.
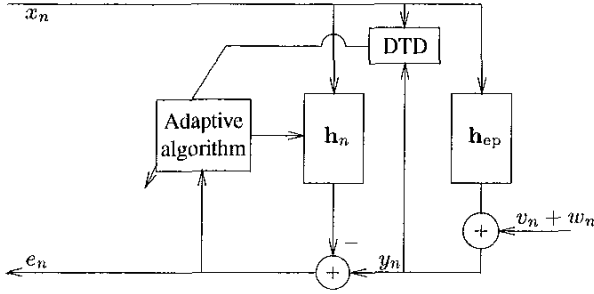
Figure 1: Block diagram of the echo canceler and double-talk detector.

The step-sizes are calculated from the last estimate of the filter coefficients so that a larger coefficient receives a larger weight, thus increasing the convergence rate of that coefficient. This has the effect that active coefficients are adjusted faster than non active coefficients (i.e. small or zero coefficients). The PNLMS algorithm is described by the following equations:

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \frac{\mu}{\mathbf{x}_n^T \mathbf{G}_n \mathbf{x}_n + \delta} \mathbf{G}_n \mathbf{x}_n e_n, \qquad (1)$$

$$\mathbf{G}_n = \operatorname{diag}\{g_{0,n}, \ldots, g_{L-1,n}\}. \qquad (2)$$

$\mathbf{G}_n$ is a diagonal matrix which adjusts the step-sizes of the individual taps of the filter, μ is the overall step-size parameter, and δ is a regularization parameter which prevents division by zero and stabilizes the solution when speech is used as input signal. The diagonal elements of $\mathbf{G}_{n+1}$ are calculated as follows, [2],

$$\gamma_{l,n+1} = \max\{\rho \max\{\delta_p, |h_{0,n}|, \ldots, |h_{L-1,n}|\}, |h_{l,n}|\}(3)$$

$$g_{l,n+1} = \gamma_{l,n+1} / \sum_{i=0}^{L-1} \gamma_{l,n+1}, \ 0 \le l \le L-1. \qquad (4)$$

Parameters δ_p and ρ are positive numbers with typical values $\delta_p = 0.01$, $\rho = 5/L$. ρ prevents coefficients from stalling when they are much smaller than the largest coefficient and δ_p regularizes the updating when *all* coefficients are zero at initialization.

A variant of this algorithm called the PNLMS++, [3], is the one we will consider here. In this algorithm, for odd-numbered time steps the matrix $\mathbf{G}_n$ is derived as above, while for even-numbered steps it is chosen to be the identity matrix ($\mathbf{G}_n = \mathbf{I}$) which results in an NLMS iteration. The alternation between NLMS and PNLMS iterations has several advantages compared to using just the PNLMS technique, e.g. it makes the PNLMS++ algorithm much less sensitive to the assumption of a sparse impulse response without sacrificing performance.

2.1. Geigel DTD

A double-talk detector is used to suppress adaptation during periods of simultaneous far- and near-end speech. A simple and efficient way of detecting double-talk is to compare the magnitude of the far-end and near-end signals and declare double-talk if the near-end magnitude becomes larger than a value set by the far-end speech. A proven algorithm that has been in commercial use for many years is the Geigel DTD, [4]. In this algorithm, double-talk is declared if

$$|y_n| \ge \vartheta \max\{|x_n|, |x_{n-1}|, \ldots, |x_{n-L+1}|\}. \qquad (5)$$

The detector threshold, ϑ, is set to 0.5 if the hybrid attenuation is assumed to be 6 dB, and to 0.71 if the attenuation is assumed to be 3 dB. A so-called hangover time, T_{hold}, is also specified such that if double-talk is detected, then the adaptation is inhibited for this duration beyond the detected end of double-talk. Although this detector works fairly well, detection errors do occur, and these result in large amounts of divergence of the adapted filter coefficients, which in turn give rise to large amounts of uncancelled echo. One way to model these large disturbances together with the background noise, is to use an outlier-contaminated stochastic process. Note that an outlier-contaminated model is not valid for the residual error in the absence of a Geigel DTD, because without the DTD the residual consists of long-lasting bursts of near-end speech. Hence the DTD is an essential component in the robust PNLMS++ algorithm to be described in the next section.

3. THE ROBUST ALGORITHMS

The NLMS and the PNLMS++ algorithm can both be made robust to large disturbances by modification of the criteria on which these algorithms are based. In general, however, such modifications lower the convergence rate. *The difficult problem of robust echo cancellation is to devise an algorithm that diverges slowly in response to double-talk, yet is able to rapidly track changes of the echo path when they occur. These two requirements are contradictory.* The key to the solution to this problem is a combination of a DTD with traditional robust statistics and a *delicately tuned scale variable, s*.

Recall that the LMS is an iterative algorithm to adjust the estimated impulse response so as to minimize the cost function, $E\{|e_n|^2\}$, i.e., the mean square error. Each iteration updates the current estimate of $\mathbf{h}_n$ by $\mu \mathbf{x}_n e_n$, which is a step in the direction of a stochastic approximation to the gradient of $E\{|e_n|^2\}$. To make the algorithm insensitive to changes of the level of input signal, x_n, the proportionate factor μ is normalized, resulting in the NLMS algorithm. It is well known, [8], that other gradient algorithms can be derived by changing the cost function to

$$J = E\{\varrho(\frac{|e_n|}{s})\} \qquad (6)$$

where $\varrho(\cdot)$ is any symmetric function with a monotonically non-decreasing derivative (with respect to its argument)[1]. s is the very important scale factor. The resulting algorithm, analogous to the steepest-descent method is

$$\mathbf{h}_n = \mathbf{h}_{n-1} - \mu \nabla_{\mathbf{h}} J. \qquad (7)$$

The algorithm can be made robust by a proper choice of J, which must be chosen such that $\lim_{e_n \to \infty} |\nabla_{\mathbf{h}} \varrho(\frac{e_n}{s})| < \infty$. Following suggestions in [7], we choose the gradient $\nabla_{\mathbf{h}} J(e_n) = E\{-\mathbf{x}_n \operatorname{sign}(e_n)\psi(\frac{|e_n|}{s})s\}$, where $\psi(\cdot)$ is a limiter,

$$\psi(\frac{|e_n|}{s}) = \min\{k_0, \frac{|e_n|}{s}\}. \qquad (8)$$

The effect of this scale factor, and the manner in which it is adapted are discussed in Section 3.1. Using a stochastic approximation of

---

[1]More generally as discussed in [7], one can use M-estimators which are defined as $J = \sum_n \varrho(\frac{|e_n|}{s})$. The choice used in (6) makes the derivation of the iterative algorithm more consistent with the derivation of the LMS algorithm.

the gradient normalizing the step-size in the same manner as in standard NLMS we get the robust NLMS algorithm,

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \frac{\mu \mathbf{x}_n}{\mathbf{x}_n^T \mathbf{x}_n + \delta} \psi(\frac{|e_n|}{s})\text{sign}\{e_n\}s. \qquad (9)$$

The PNLMS algorithm given in (1) can be made robust in an exactly analogous manner, yielding the update equation,

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \frac{\mu \mathbf{G}_n \mathbf{x}_n}{\mathbf{x}_n^T \mathbf{G}_n \mathbf{x}_n + \delta} \psi(\frac{|e_n|}{s})\text{sign}\{e_n\}s. \qquad (10)$$

Alternating the iterations with $\mathbf{G}_n$ as given in (3) and the identity matrix then yields the robust PNLMS++ algorithm.

3.1. Estimating the scale factor

The estimate of the scale factor, s, should reflect the background noise level at the near-end, be robust to short burst disturbances (double-talk) and track long term changes of the residual error (echo path changes). To fulfill these requirements we have chosen the scale factor estimate as,

$$s_n = \lambda s_{n-1} + \frac{1-\lambda}{\beta} s_{n-1} \psi(\frac{|e_n|}{s_{n-1}}), \qquad (11)$$

where $s_{-1} = \sigma_x$. The choice of this method of estimating s is justified in [9]. With this choice, the current estimate of s is governed by the level of the error signal in the immediate past over a time interval roughly equal to $1/(1 - \lambda)$. When the algorithm has not yet converged, s is large. Hence the limiter is in its linear portion and therefore the robust algorithm behaves like the conventional NLMS or PNLMS algorithms. When double-talk occurs, the error is determined by the limiter and by the scale of the error signal during the recent past of the error signal *before* the double-talk occurs. Thus divergence rate is reduced for a duration of about $1/(1 - \lambda)$. This gives ample time for the DTD to act. If there is a system change, the algorithm will not track immediately. However, as the scale estimator tracks the larger error signal the nonlinearity is scaled up and convergence rate accelerates. The trade off between robustness and tracking rate of the adaptive algorithm is thus governed by the tracking rate of the scale estimator which is controlled by one single parameter, λ. As with the Geigel DTD, it is useful to introduce a hangover time for control of scale updating. When the DTD detects double-talk, adaptation of s_n should be inhibited for a specific time, preferable as long as the DTD hangover time, T_{hold}.

3.2. Generalization of PNLMS to the Affine Projection Algorithm

Let $\mathbf{y}_n = [y_n \ \ldots \ y_{n-p+1}]^T$, be a vector of samples y_n and $\mathbf{X}_n = [\mathbf{x}_n, \ldots, \mathbf{x}_{n-p+1}]$ the excitation matrix where p is the projection order. A residual echo vector $\mathbf{e}_n = \mathbf{y}_n - \mathbf{X}_n^T \mathbf{h}_n$, A proportionate affine projection based algorithm (PAPA) is then given by,

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mu \mathbf{G}_n \mathbf{X}_n (\mathbf{X}_n^T \mathbf{G}_n \mathbf{X}_n + \delta \mathbf{I})^{-1} \mathbf{e}_n, \qquad (12)$$

where $\mathbf{G}_n$ is as defined in the Section 2 and $\mathbf{R}_{xx,n}^{-1} = (\mathbf{X}_n^T \mathbf{G}_n \mathbf{X}_n + \delta \mathbf{I})^{-1}$ is a weighted estimate of the inverse correlation matrix of the input signal. This matrix "whitens" the input data, $\mathbf{X}_n$, and thus the convergence rate of the adaptive filter is increased. With $\mathbf{G}_n = \mathbf{I}$ and $\delta = 0$, equation (12) reduces to the standard APA, first introduced in [6]. The regularization parameter in (12) was

proposed in [10]. As evident, PAPA is obtained by combining APA with the proportionate step-size of PNLMS. We can omit the matrix $\mathbf{G}_n$ in the definition of $\mathbf{R}_{xx,n}$ to save computations. Inclusion of the matrix requires a significant $p^2 L$ multiplications per sample, but according to our simulations, the effect on performance and stability is minimal.

A robust version of PAPA (and hence of APA) is obtained straightforwardly, by applying the principles presented previously:

$$\mathbf{h}_n = \mathbf{h}_{n-1} + \mu \mathbf{G}_n \mathbf{X}_n \mathbf{R}_{xx,n}^{-1} \psi(\mathbf{e}_n). \qquad (13)$$

$\psi(\mathbf{e}_n) = [\min\{\frac{|e_n|}{s_{n-1}}, k_0\} \odot \text{sign}(\mathbf{e}_n)]s_{n-1}$ where $\odot$ denotes elementwise multiplications and $|\cdot|$ in $|\mathbf{e}_n|$ operates on the individual elements.

Additionally most of the computational procedures of the Fast Affine Projection (FAP) algorithm, [11, 12], can be incorporated in order to reduce the computational complexity of PAPA.

4. SIMULATIONS

The purpose of these simulations is to show the performance of the robust algorithms during double-talk and the high convergence rate of the robust PAPA and PNLMS algorithms when speech is used as excitation signal. The order p is chosen as 2 in order to compromise between complexity and performance.

The parameter settings chosen for the simulations are: $\mu = 0.2$, $L = 512$ (64 ms), $\delta = 2 \cdot 10^5$ (NLMS, PNLMS++), $\delta = 1 \cdot 10^6$ (PAPA), $\delta_p = 0.01$, $\rho = 0.01$. $\sigma_x = 1900$, SNR= 39 dB (echo-to-noise ratio). Average far-end to double-talk ratio is 6 dB. The hybrid attenuation is 20 dB and the Geigel detector assumes 6 dB attenuation. Parameters for the robust algorithm are, $(\lambda, k_0) = (0.997, 1.1)$. This choice results in $\beta \approx 0.60665$. $\mathbf{h}_{-1} = \mathbf{0}$. $s_{-1} = 1000$. The scale estimate in (11), s_n, is never allowed to become lower than 2. This inhibits bad behavior in low noise situations. All algorithms are tuned to achieve the same minimum misalignment in order to fairly compare convergence rate. The misalignment is given by, $\varepsilon = \|\mathbf{h} - \mathbf{h}_{\text{ep}}\| / \|\mathbf{h}_{\text{ep}}\|$ where $\mathbf{h}_{\text{ep}}$ is the true echo path. The impulse response and corresponding magnitude function of the hybrid is shown in Fig. 2a, b. Figure 3 shows far-end signal, double-talk and the misalignment of the three algorithms. Initial convergence rates of PNLMS++ and PAPA are clearly superior to that of NLMS. While the non-robust NLMS (with Geigel detector!) diverges to a misalignment of +5 dB the robust algorithms are much less affected and never perform worse than −10 dB misalignment during double-talk. The slow reconvergence of NLMS and robust PNLMS++ after the double-talk sequence is caused by poor excitation of the speech signal, i.e. this segment is highly correlated.

Figure 4 shows the behavior after an abrupt system change where the impulse response is shifted 200 samples at 1 second. The robust algorithms outperform NLMS in this case.

5. CONCLUSIONS

A scaled nonlinearity combined with a Geigel DTD increases the robustness of the echo canceler. The scaled nonlinearity operates in the same manner as varying the step-size (μ), i.e., bounding the error signal can be interpreted as a reduction of the step-size parameter. What differentiates the approaches is that traditional variable step-size methods try to detect periods of double-talk and then take action, while the robust technique uses the signal before
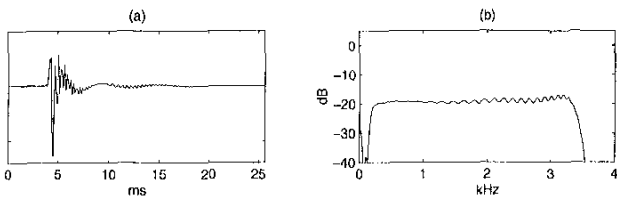
Figure 2: Impulse response (a) and magnitude (b) respectively of the frequency response of the hybrid in the simulation. Hybrid attenuation: 20 dB.

Figure 4: Performance for speech signal after abrupt hybrid changes occurring at 1 s. Other conditions same as in Fig. 3.
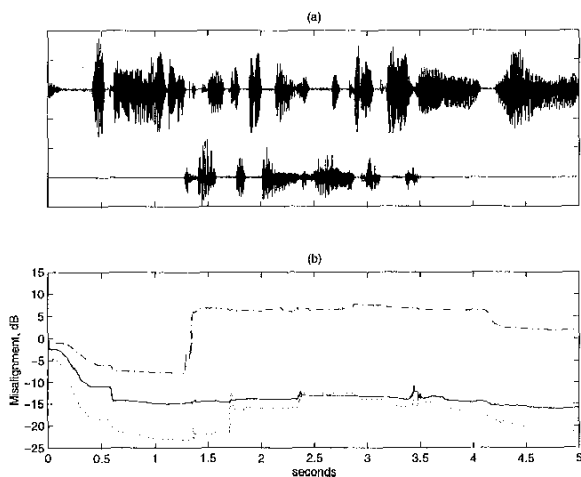

Figure 3: Performance for speech signal during double-talk. Far-end (upper) and near-end (lower) signals. Average far- to near-end ratio: 6 dB (1.125-3.625 s) (a). Misalignment (b), $\lambda = 0.997$, $k_0 = 1.1$, hybrid attenuation: 20 dB. Solid line: Robust PNLMS++, Dashed line: NLMS, Dotted line: Robust PAPA.

double-talk in order to be prepared for it. Due to this fact the robust technique is faster and more efficient. Another major advantage is that only a few instructions and little memory is required to implement the robust principle.

The complexity of PNLMS and PAPA is in the order of $4L$ multiplications per sample which is about twice that of NLMS. However, the convergence rate of PNLMS and PAPA is considerably higher. It is shown in this paper that the robust version of PNLMS and PAPA converge faster than NLMS and perform significantly better. Simulations not shown here confirm reliable performance of the robust algorithms for different double-talk situations. The principle of robustness works at all stages of convergence of the robust algorithms. They resist divergence during double-talk even though they have not yet fully converged. It should also be mentioned, that the performance loss using the nonlinearity for robustness is only minor.
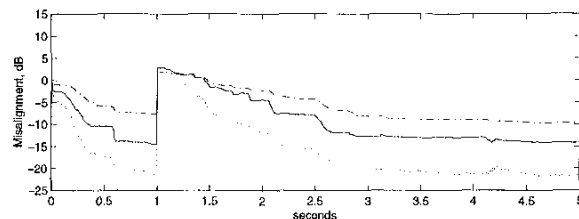
Acknowledgments

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, ch. 9, Prentice Hall, 1996.

[2] D. L. Duttweiler, "Proportionate normalized least mean squares adaptation in echo cancelers," *IEEE Trans. Speech Audio Processing*, Submitted.

[3] S. L. Gay, "An efficient, fast converging adaptive filter for network echo cancellation," *Proc. of Assilomar*, Nov., 1998.

[4] D. L. Duttweiler, "A twelve-channel digital echo canceler," *IEEE Trans. Commun.*, Vol. 26 , pp. 647-653, May 1978.

[5] T. Gänsler, "A double-talk resistant subband echo canceller," *Signal Processing*, Vol. 65, 1998.

[6] K. Ozeki and T. Umeda, "An adaptive filtering algorithm using an orthogonal projection to an affine subspace and its properties," *Electronics and Communications in Japan*, Vol. 67-A, no. 5, 1984.

[7] P. J. Huber, *Robust Statistics*, pp 68-71, 135-138, 109, Wiley, New York, 1981.

[8] M. M. Sondhi, "An adaptive echo canceler," *Bell System Technical Journal*, XLVI vol. 3, pp. 497-510, March. 1967.

[9] T. Gänsler, Steven Gay, M. Mohan Sondhi, and J. Benesty "Double-talk robust fast converging algorithms for network echo cancellation," *IEEE Trans. Speech Audio Processing*, Submitted.

[10] S. G. Kratzer and D. R. Morgan, "The partial-rank algorithm for adaptive beamforming," *Proc. SPIE Int. Soc. Optic. Eng.*, Vol. 564, pp. 9-14, 1985.

[11] S. L. Gay and S. Tavathia, "The fast affine projection algorithm," Proc. IEEE ICASSP, pp. 3023-3026, 1995.

[12] M. Tanaka, Y. Kaneda, S. Makino, and J. Kojima, "Fast projection algorithm and its step size control," Proc. IEEE ICASSP, pp. 945-948, 1995.