Iterative Speech Enhancement With Spectral Constraints

John H. Hansen and Mark A. Clements

Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

Abstract

A new and improved iterative speech enhancement technique based on spectral constraints is presented in this paper. The iterative technique, originally formulated by Lim and Oppenheim, attempts to solve for the maximum likelihood estimate of a speech waveform in additive white noise. The new approach applies inter- and intra-frame spectral constraints to ensure convergence to reasonable values and hence improve speech quality. An extremely efficient technique for applying these constraints is in the use of line spectral pair (LSP) coefficients. The inter-frame constraints ensures more speech-like formant trajectories than those found in the unconstrained approach. Results from speech degraded by additive white Gaussian noise show noticeable quality improvement.

Introduction

The successfulness of an enhancement algorithm rests on the goals and assumptions used in deriving the approach. Depending on the application, a system may be directed at one or more objectives such as improving overall quality, increasing intelligibility, reducing listener fatigue, etc. Three assumptions normally made include: i) that the noise distortion be additive, ii) that only the degraded speech signal is available, and iii) that the noise and speech signals are uncorrelated. In general, constraints placed on the speech model improve the potential for separating speech from background noise. However, such systems are also more sensitive to "deviations" from these constraints. The degradation considered is additive white Gaussian noise. The basis of the technique is an iterative enhancement approach based on noncausal Wiener filtering originally formulated by Lim and Oppenheim [1]. This approach attempts to solve for the maximum likelihood estimate of a speech waveform in additive white noise using the constraint that the signal is an all-pole process. Crucial to the success of this approach is the accuracy of the estimates of the all-pole speech parameters at each iteration. One advantage of the Wiener filtering approach is that no "musical tone" artifacts are present after processing as can be observed in spectral subtraction techniques. In addition, under certain conditions, it can be shown that it is the optimal solution in the mean-squared sense for a white noise distortion. Although successful in a mathematical sense, this technique has received little application due to several factors. First, it is an iterative scheme with sizable computational requirements as opposed to a direct form such as spectral subtraction. Second, although the original sequential MAP estimation technique was shown to increase the joint likelihood of the speech waveform and all-pole parameters, heuristic convergence criteria had to be employed. After an extensive investigation [2], this approach was found to produce significant levels of enhancement for white Gaussian noise in 3-4 iterations. The technique was generalized to allow for colored aircraft noise. Various spectral estimation techniques where employed for securing estimates of the colored background noise and although the noise was not stationary, estimates were performed prior to application of the algorithm. With these assumptions, good enhancement took place in 2-3 iterations. It is assumed that in a real-time environment however, noise spectral estimates could be gathered and updated during silent intervals. An important observation which could be made from this previous work was that as additional iterations were performed, individual formants of the speech decreased in bandwidth (see fig.1), resulting in unnatural sounding speech. Frame-to-frame pole jitter was also observed which contributed to unnatural sounding results. Also, the original technique employs no explicit frame-to-frame constraints. Since the original algorithm already constrains the speech to be the response from an all-pole system, applying further constraints on the pole movements may improve the algorithms performance. One set of constraints were applied directly to the LPC poles. These results were quite encouraging, yet computationally intensive. A new approach for implementing the spectral constraints was formed by employing the line spectral pair (LSP) transformation as a method for representing the vocal tract spectrum. This method of specification allowed constraints to be efficiently applied to the speech model pole movements across time (inter-frame) so that formants lay on smooth tracks. In addition, constraints could also be easily applied across iterations (intra-frame) on a frame-by-frame basis.

Iterative Speech Enhancement

Enhancement based on the estimation of all-pole speech parameters in additive white Gaussian noise was investigated by Lim and Oppenheim [1], and later for a colored noise degradation by Hansen and Clements [2]. It was shown that the estimation procedures which result in linear equations without background noise, become nonlinear when noise is introduced. However by allowing a suboptimal procedure, an iterative algorithm results which possesses the property that the estimation procedure is linear at each iteration.

Consider the statistical parameter estimation of speech in the presence of noise. Over a short-time basis, the speech signal can be represented as the following difference equation:

$$s(n) = \mathbf{a}^T s(n-1,n-p) + g\,w(n) \qquad (1)$$

where $\mathbf{a}^T=[a_1,a_2,\ldots,a_p]$ represents the all-pole predictor coefficients. Substituting the degraded speech into the speech model gives the following equation for the observation vector:

$$\mathbf{Y_0} = y(N-1,0) = s(N-1,0) + d(N-1,0) \qquad (2)$$
$$\mathbf{Y_0} = \mathbf{a}^T y(n-1,n-p) + g\,w(n) + d(n) - \mathbf{a}^T d(n-1,n-p)$$

where s(N-1,0) are N samples of original speech, and d(N-1,0) represents the additive background noise. The 2p + 1 unknowns include the predictor coefficients **a**, initial conditions for the predictor given by $\mathbf{S_i} = s(-1,-p)$, and the gain factor g for the input excitation. Consider the case where all unknown parameters are random with a priori Gaussian probability density functions. The basic procedure used is a maximum a priori (MAP) estimator, which maximizes the probability density function of

CH2396-0/87/0000-0189 $1.00 © 1987 IEEE

the parameters given the observations. Therefore, a,g,S_i are chosen to maximize the probability density function $p(\mathbf{a},g,\mathbf{S_i}|\mathbf{Y_0})$. The procedure requires that **a** be chosen to maximize $p(\mathbf{a}|\mathbf{Y_0})$, noting that the estimate is conditioned on the noisy observations $\mathbf{Y_0}$. Using Bayes' rule, $p(a|\mathbf{Y_0})$ can be written as a product of terms involving $p(\mathbf{Y_0}|\mathbf{a},g,\mathbf{S_i})$. When the Gaussian density function $p(\mathbf{Y_0}|\mathbf{a},g,\mathbf{S_i})$ is expanded, it can be shown that the mean and variance are functions of the predictor coefficients **a**. Therefore the resulting equations for maximizing $p(a|\mathbf{Y_0})$ are nonlinear, involving partial derivatives with respect to **a**. Lim and Oppenheim considered a suboptimal solution employing a two step approach based on MAP estimation of $\mathbf{S_0}$ given $\mathbf{Y_0}$, followed by MAP estimation of **a** given $\hat{\mathbf{S}}_0$,where $\hat{\mathbf{S}}_0$ is the result of the first estimation. Observations indicate that this algorithm converges to a local maximum of the joint density $p(\mathbf{a},\mathbf{S_0}|\mathbf{Y_0};g,\mathbf{S_i})$. In particular, if the probability density function is unimodal, and the initial estimate for **a** is such that the local maximum equals the global maximum, then the procedure is equivalent to the joint MAP estimate of **a** and $\mathbf{S_0}$. After some simplification, the MAP estimation of $\mathbf{S_0}$, based on maximizing the probability density function $p(\mathbf{S_0}|\mathbf{a_i},\mathbf{Y_0})$ which is jointly Gaussian in $\mathbf{Y_0}$, is equivalent to a minimum mean squared error (MMSE) estimate of $\mathbf{S_0}$. Therefore as the observation window increases in length, the procedure for obtaining a MMSE estimate of s(n) approaches a noncausal Wiener filter. With this, the implementation of the algorithm is presented in Figure 2. This approach can also be extended to the colored noise case as shown. As indicated, the background noise spectral density must be estimated during non-speech activity.

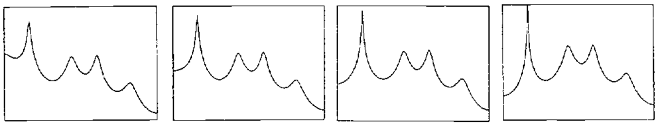
(a) Distorted Original *(b) 2 Iterations* *(c) 4 Iterations* *(d) 8 Iterations*

Figure 1: Variation in vocal tract response across iterations.

As indicated, the sequential MAP estimation technique increases the joint likelihood of the speech waveform and all-pole parameters, yet a heuristic convergence criterion had to be employed. Also, as additional iterations were performed, individual formants of the speech decrease in bandwidth as indicated in figure 1. Frame-to-frame pole jitter was also observed. Both effects contributed to unnatural sounding speech. The goal, therefore is to impose constraints on the pole movements across time (inter-frame) and iterations (intra-frame). An initial approach was to limit the poles from moving too close to the unit circle by performing an off-axis spectral evaluation where the z-transform is evaluated on a circle further away from the poles of the spectral model. Other approaches considered included applying constraints directly to the pole radii and/or angular displacements in the LPC model. Performance of such inter and intra-frame constraints lead to encouraging results, but at the expense of a pth order root-solve and a pole ordering step per frame for each iteration. Since root solving is not always numerically accurate and ordering can be inconsistent across frames, a more robust approach was sought to implement these constraints. Previous success of the line spectral pair (LSP) transformation in speech coding by Crosmer [3], led to the use of LSP's for this purpose.

Line Spectral Pair Representation of Spectral Characteristics

The LSP transformation may be viewed as an alternative representation of the LPC spectrum. The LSP coefficients are obtained from the LPC prediction coefficients by combining the forward and backward predictor polynomials as follows:

$$P(z) = A(z) + B(z), \qquad Q(z) = A(z) - B(z). \qquad (3)$$

The vocal tract transfer function is given by $g/A(z)$, and M is the order of the LPC speech model. The resulting polynomials P(z) and Q(z), are symmetric and antisymmetric, respectively, with a root of P(z) at $z=+1$, and a root of Q(z) at $z=-1$. The remainder of the roots of P and Q all lie on the unit circle. Since the roots occur in conjugate pairs, the original polynomial can be represented by M real numbers. The angles of the roots, $\{\omega_i,\ i=1,2,\ldots,M\}$, are called the *line spectrum pairs*.

The LSP's possess several important properties which make them attractive for use in applying spectral constraints. One important characteristic is that if the vocal tract polynomial A(z) has all its roots inside the unit circle (i.e., a stable filter), then the roots of P and Q will be interleaved around the unit circle [3]. If two adjacent LSP frequencies are identical, it indicates that a root of A(z) lies on the unit circle.

In addition to their attractive representation of the LPC spectrum, the LSP coefficients offer the possibility of a more direct representation of perceptually important information. Specifically, their is a firm statistical relationship between the locations and bandwidths of the speech formants and the locations of the roots of P and Q respectively. Since roots of the P polynomial correspond approximately to locations of formant center frequencies (when a formant is present), the P polynomials' LSP coefficients are termed *position coefficients*. It can be shown that the closer two LSP coefficients are together, the narrower the bandwidth of the corresponding pole of the vocal tract filter. Therefore, formants are indicated when two LSP coefficients are close together. When LSP coefficients are far apart, they indicate poles which contribute only to the overall spectral shape. Because of their relationship to the presence or absence of a formant by their nearness to a position coefficient, the coefficients of Q are termed *difference coefficients*. Given the LSP coefficients, the position coefficients are simply the odd index LSP coefficients, $\{p_i=\omega_{2i-1},\ i=1,2,\ldots,M/2\}$. The difference coefficients are given as follows:

$$\{|\,d_i\,| = \underset{j=-1,1}{\mathrm{MIN}}\ (\,|\,\omega_{2i+j} - \omega_{2i}\,|\,),\ i = 1,2,\ldots,M/2\} \qquad (4)$$

where the sign of d_i is positive if ω_{2i} is closer to ω_{2i+j}, and otherwise is negative. With this interpretation, a new enhancement technique based on Wiener filtering is now possible by imposing constraints on the LSP coefficients.

Step 1: Estimate $\mathbf{a}_i$ from $S_{O,i}$.
Use either: i. first P values as the initial condition vector
or: ii. always assume $\mathbf{S_i} = \mathbf{0}^T$.

Step 2: i. Using $\hat{\mathbf{a}}_i$, estimate the speech spectrum:

$$P_S(\omega) = \frac{g^2}{\left|1 - \sum_{k=1}^{P} a_k e^{-jk\omega}\right|^2}$$

ii. Calculate gain term using Parseval's theorem.
iii. Estimate either the degrading
a.) white noise variance σ_d^2, or **b.)** colored noise spectrum $P_D(\omega)$ from a period of silence closest to the utterance.
iv. Construct the noncausal Wiener filter;

$$\textbf{a.)}\ H(\omega) = \frac{P_S(\omega)}{P_S(\omega) + \sigma_d^2} \qquad \textbf{b.)}\ H(\omega) = \frac{P_S(\omega)}{P_S(\omega) + P_D(\omega)}$$

v. Filter the estimated speech $\hat{s}_i$ to produce $\hat{s}_{i+1}$.
vi. Repeat until some specified error criterion is satisfied, $\Delta\epsilon <$ THRESHOLD.

Figure 2: Enhancement Algorithm based on All-pole modeling/Wiener filtering. **a)** a AWGN distortion **b)** a non-white distortion

Enhancement with Spectral Constraints

Consider the statistical parameter estimation of speech in the presence of noise, where all unknown parameters are random with a priori Gaussian probability density functions. It can be shown that MAP estimation of **a**, g, and $\mathbf{S_i}$ given the noisy observations $\mathbf{Y_0}$, results in a set of nonlinear equations. Therefore, instead of joint estimation of **a** and $\mathbf{S_0}$, a suboptimal solution is formulated employing a two step approach based on

MAP estimation of $\mathbf{S_0}$ given $\mathbf{Y_0}$, followed by MAP estimation of $\mathbf{a}$ given $\hat{\mathbf{S}}_\mathbf{0}$, where $\hat{\mathbf{S}}_\mathbf{0}$ is the result of the first estimation. Since speech can be considered short-time stationary, frame-to-frame spectral constraints may aid in enhancement. The new approach imposes such constraints on the vocal tract spectrum between MAP estimation steps. The procedure for obtaining the MAP estimate of $\mathbf{a}$ from MAX $p(\mathbf{a}|\hat{\mathbf{S}}_\mathbf{0};g,\mathbf{S_I})$ remains the same. The next step is to apply spectral constraints to $\hat{\mathbf{a}}_i$ which will ensure that; i) the all-pole speech model is stable, ii) it possess speech-like characteristics (i.e., poles are not too close to the unit circle causing narrow bandwidths), and iii) the vocal tract characteristics do not vary wildly from frame-to-frame when speech is present. Due to this constrained approach, an improved estimate $\hat{\mathbf{a}}_i'$ results. Given this new estimate, the second MAP estimation of $\mathbf{S_0}$ given $\hat{\mathbf{a}}_i'$ can be carried out by maximizing $p(\mathbf{S_0}|\hat{\mathbf{a}}_i',\mathbf{Y_0};g,\mathbf{S_I})$. Since $p(\mathbf{S_0}|\hat{\mathbf{a}}_i',\mathbf{Y_0};g,\mathbf{S_I})$ is still jointly Gaussian in $\mathbf{Y_0}$, the resulting MAP estimate is equivalent to a MMSE estimate of $\mathbf{S_0}$. Again, in the limiting case, the procedure for obtaining the MMSE estimate of s(n) approaches a noncausal Wiener filter. Once this new estimate of $\hat{\mathbf{S}}_{\mathbf{0},i}$ is formed, the iterative procedure continues by re-estimating $\hat{\mathbf{a}}_i$, applying constraints to $\hat{\mathbf{a}}_i$, and then forming the noncausal filter using $\hat{\mathbf{a}}_i'$ to re-estimate $\hat{\mathbf{S}}_{\mathbf{0},i}$. This continues until some convergence criterion is satisfied. The procedure for implementing these constraints will now be addressed.

Two classes of spectral constraints are considered; inter-frame (across time), and intra-frame (across iterations). Two approaches are considered: a fixed frame rate, and a variable frame rate approach. In the first of these, the LPC predictor coefficients, $\mathbf{a}$, are first converted to LSP position and difference coefficients. Next, each frame's energy is observed, and if it is above some threshold, it is classified as voiced speech; if it is below, then it is either noise or unvoiced speech. A local running count L_1, is kept for the number of consecutive frames which fall below the energy threshold. If L_1 reaches L_{MAX}, then all subsequent frames below the threshold are classified as noise. This allows for further smoothing for long periods of silence. The position coefficients for each frame are smoothed using a weighted triangular window with a variable base of support (1 to 5 frames). If a frame has been classified as noise, maximum smoothing is performed. In addition, the lower formant frequencies are smoothed over a narrower triangle width than for those position coefficients at higher frequencies. This preserves perceptually important speech characteristics found in the lower formants. No smoothing is performed on the difference coefficients since they are more closely related to formant bandwidth than formant location. However, it is possible that a difference coefficient falls within a "forbidden zone," (i.e., the region within d_{MIN} of a position coefficient). When this occurs, the LPC analysis has most likely overestimated the Q of a particular pole. Since this causes unnatural sounding speech, (as in the unconstrained approach), the value of $|d_i|$ is set to d_{MIN}. Finally, the position and difference coefficients are combined to form the constrained LPC predictor coefficients $\hat{\mathbf{a}}_i'$.

The second inter-frame constraint approach considered is a variable frame rate technique which takes advantage of the interpolation properties of the LSP coefficients. The speech signal is first divided into segments, where segments are chosen such that they are long when the speech spectrum is varying slowly and short when the speech spectrum is varying quickly. The LSP coefficients are reconstructed with linear interpolation used to compute the coefficients for intermediate frames.

The segmentation algorithm begins with a step to determine the onset/offset of speech. This is carried out by thresholding the LPC residual energy, which produces relatively long segments. Next, the long segments are subdivided based on the curvature of the position coefficients. This is performed by computing a gain-normalized Itakura-Saito measure of the spectral distance between the frequency response of two adjacent frames. The procedure continues by computing the distortion of position coefficients for successively longer segments until the distortion exceeds a threshold T_D. At that point, a subsegment boundary is set, with the intermediate position coefficients reconstructed via linear interpolation. During this step, the length of a subsegment is also limited to L_{MAX} to prevent excessively long segments which might contribute to muffled or unnatural sounding speech. The advantage of this approach is that it incorporates more information from adjacent frames when the spectrum indicates similar characteristics. Yet, it also reduces the effects of adjacent frames when the spectrum is significantly different as in the case of a transition from unvoiced passages to noise. This in effect, distorts the position coefficients as little as possible when associated difference coefficients indicate the presence of formants. Difference coefficients for each frame, (or an average set across a segment) are used to compute the predictor coefficients $\hat{\mathbf{a}}_i'$. The difference coefficients are required to be at least d_{MIN} or greater in distance from adjacent position coefficients to ensure that poles from the LPC filter do not move too close to the unit circle.

Inter-frame constraints are applied to a single frame across iterations, and as such require the frames' previous estimates to be available. The motivation for such constraints is that under certain conditions, pole locations for the same frame vary significantly from their previous estimated values. Since the present estimate of $\hat{\mathbf{a}}_i$ affects the next estimate of $\hat{\mathbf{S}}_{\mathbf{0},i}$, sections of $\hat{\mathbf{S}}_{\mathbf{0}_i}$ will also vary significantly across iterations. In addition, previous results based on objective speech quality measures indicated that the unconstrained approach produced minimum objective measures at different iterations for different classes of speech. For example, maximum overall speech quality was observed for additive white Gaussian noise in three iterations. This was also true for vowels and fricatives. However, glides required two iterations, nasals, liquids, and affricates between five and six. It is therefore desirable to be able to affect the convergence rate so that the best objective measure of quality occurs at the same iteration across all classes of speech. Improved quality as measured by objective measures may also result in improved estimation of $\hat{\mathbf{a}}_i$. By constraining the vocal tract filter to be a function of its previous estimates, it may be possible to accomplish this. Two approaches are considered, one applied to the autocorrelation lags, the other to the position coefficients. The first approach simply weights the present set of autocorrelation lags with the same frame from previous iterations. This technique is very easy to perform, since the autocorrelation lags must be computed in order to estimate the predictor coefficients $\mathbf{a}$. The second approach weights position coefficients with those from the same frame but previous iteration. If the corresponding difference coefficient indicates the adjacent position coefficient to represent a formant, this approach has the effect of constraining the formants to lie along smooth tracks across iterations.

Results

Speech degraded by additive white Gaussian noise was processed using various configurations of the new constrained enhancement algorithm. Energy thresholds for inter-frame constraints were obtained from frame energy histograms at each signal-to-noise ratio. Excellent enhancement resulted for a wide range of threshold values. Intra-frame constraints were applied across two to three iterations. Informal listening tests indicated noticeable quality improvement, although no intelligibility testing has been performed. However, there has been extensive work carried out in the area of objective speech quality measures [4]. Good correlation has been shown to exist between subjective quality and objective measures. Therefore, objective measures including: the Itakura-Saito likelihood ratio, log area ratio, and weighted spectral slope measure where used for evaluation. Figure 3 illustrates a comparison of typical results for the various constraint approaches. Itakura-Saito measure is plotted versus signal-to-noise ratio for a white noise distortion. Plot *a* represents the original distorted speech. Plots *b* through *e* represent combinations of inter-frame constraints (both fixed and variable rate), and intra-frame constraints (applied to position coefficients/autocorrelation lags). All configurations examined showed significant improvement in Itakura-Saito measures. Threshold settings for the variable frame rate inter-frame constraint were somewhat sensitive to varying noise levels. However, the fixed frame approach by itself, and with either autocorrelation or position intra-frame constraints gave impressive results with little sensitivity to varying levels of SNR. In order to determine a limit on the level of enhancement, the original undistorted predictor coefficients **a** were used in the unconstrained algorithm. In essence, the two step MAP estimation approach is now reduced to a single MAP estimate of $\mathbf{S_0}$, and therefore represents the theoretical limit for enhancement using Wiener filtering. Plot *f* indicates this limit. Although only Itakura-Saito measures are shown, similar improvement was also observed for log area ratios and weighted spectral slope measures. Figure 4 compares the new approach to existing techniques. Plot *b* shows results from spectral subtraction as formulated by Boll [5]. An evaluation was performed for both half and full-wave rectification, along with one to five frames of magnitude averaging; where these points represent the best results. Plot *c* is from the unconstrained Wiener filtering technique. Plots *d* and *e* are typical values for the inter-frame constraint (fixed frame rate), and inter plus intra-frame constraints (fixed frame and autocorrelation lags). Again *f* indicates the limit for the Wiener filtering approaches.

Sound	*Itakura-Saito Likelihood Measure*			
Type	*Original*	*Lim-Oppenheim*	*Hansen-Clements*	*True LPC*
Silence	1.634	1.649	0.842	0.319
Vowel	4.020	3.299	1.651	0.582
Nasal	19.814	17.656	3.968	0.324
Stop	7.261	3.979	1.099	0.435
Fricative	3.739	3.509	1.766	0.649
Glide	1.525	1.442	1.131	0.705
Liquid	9.597	4.545	0.998	0.303
Affricate	3.924	2.702	2.229	0.323
Voiced + Unvoiced	5.838	4.293	1.761	0.519
Total	4.022	3.151	1.364	0.433

SNR=+5dB

Table 1: Comparison of algorithms over sound types for white Gaussian noise.

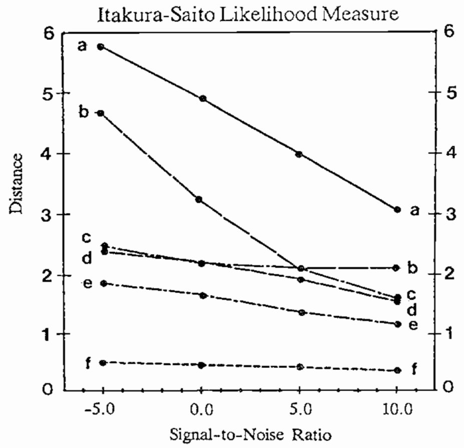

Figure 3: Comparison of constraint algorithms over SNR.
a.) Original Distorted Speech
b.) Inter-Frame Constraint: Variable Frame
c.) Inter-Frame Constraint: Fixed Frame
d.) Inter & Intra-Frame Constraints: Fixed Frame, Position
e.) Inter & Intra-Frame Constraints: Fixed Frame, Autocorrelation
f.) Theoretical limit: using undistorted LPC coefficients, **a**.

Performance evaluation over sound classes was accomplished by hand partitioning speech into segments. Entire sentences were processed, and objective measures from each class were computed. Table 1 summarizes this comparison between the unconstrained Lim-Oppenheim technique to that of the inter and intra-frame constraint approach. Measures for the theoretical limit using undistorted LPC predictor coefficients **a** are also indicated. Improvement is indicated for all types of speech. In addition, the constrained approach produced superior objective measures of quality across all speech classes at the same iteration. These results clearly indicate improvement over the unconstrained approach as well as spectral subtraction for additive white Gaussian noise.

Conclusions

The application of spectral constraints to noncausal Wiener filtering results in improved speech enhancement. Informal listening tests along with objective measures such as Itakura-Saito and log-area-ratio's show improvement over the unconstrained technique. By using the Line Spectral Pair transformation, a modest increase in computational requirements results in significant improvement in speech quality. This approach to pole movement constraints is quite robust over direct methods applied to pole radial/angular movements. Finally, this approach may be useful in enhancement for human listeners as well as a preprocessor for speech recognition.

References

[1] J.S. Lim, A.V. Oppenheim, " All-Pole Modeling of Degraded Speech," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-26, pp. 197-210, June 1978.

[2] J.H. Hansen, M.A. Clements, " Enhancement of Speech Degraded By Non-White Additive Noise," Technical Report DSPL-85-6, Georgia Institute of Technology, Atlanta, August 1985.

[3] J.R. Crosmer, " Very Low Bit Rate Speech Coding Using the Line Spectrum Pair Transformation of the LPC Coefficients, " Ph.D. dissertation, School of Electrical Engineering, Georgia Institute of Technology, Atlanta, June 1985.

[4] S.R. Quackenbush, " Objective Measures of Speech Quality, " Ph.D. dissertation, School of Electrical Engineering, Georgia Institute of Technology, Atlanta, May 1985.

[5] S.F. Boll, " Suppression of Acoustic Noise in Speech using Spectral Subtraction, " *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-27, pp. 113-120, April 1978.

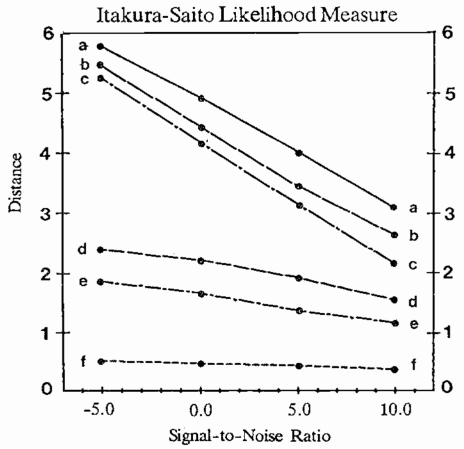

Figure 4: Comparison of enhancement algorithms over SNR.
a.) Original Distorted Speech
b.) Boll: Spectral Subtraction, using magnitude averaging
c.) Lim-Oppenheim: Unconstrained Wiener filtering
d.) Hansen-Clements: employing Inter-Frame constraints
e.) Hansen-Clements: employing Inter & Intra-Frame constraints
f.) Theoretical limit: using undistorted LPC coefficients, **a**